UNIVERSITY OF
LINCOLN

Automatic Classification of Flying Bird Species using Computer Vision Techniques

J Atanbori

Doctor of Philosophy

2017

AUTOMATIC CLASSIFICATION OF FLYING BIRD SPECIES USING COMPUTER VISION TECHNIQUES

John Atanbori

A thesis submitted in partial fulfilment of the requirements of
the University of Lincoln for the degree of Doctor of Philosophy

June 2017

Declaration

I declare that no portion of the work contained in this thesis has been submitted in support of an application for a degree or qualification of this or any other university or other institution of learning. All verbatim extracts have been distinguished by quotation marks, and all sources of information have been specifically acknowledged. This research has been documented, in part, within publications listed in Chapter 1 of this thesis.

John Atanbori

Date: June 2017

Abstract

Bird species are recognised as important biodiversity indicators: they are responsive to changes in sensitive ecosystems, whilst populations-level changes in behaviour are both visible and quantifiable. They are monitored by ecologists to determine factors causing population fluctuation and to help conserve and manage threatened and endangered species. Every five years, the health of bird population found in the UK are reviewed based on data collected from various surveys.

Currently, techniques used in surveying species include manual counting, Bioacoustics and computer vision. The latter is still under development by researchers. Hitherto, no computer vision technique has fully been deployed in the field for counting species as these techniques use high-quality and detailed images of stationary birds, which make them impractical for deployment in the field, as most species in the field are in-flight and sometimes distant from the cameras field of view. Techniques such as manual and bioacoustics are the most frequently used but they can also become impractical, particularly when counting densely populated migratory species. Manual techniques are labour intensive whilst bioacoustics may be unusable when deployed for species that emit little or no sound.

There is the need for automated systems for identifying species using computer vision and machine learning techniques, specifically for surveying densely populated migratory species. However, currently, most systems are not fully automated and use only appearance-based features for identification of species. Moreover, in the field, appearance-based features like colour may fade at a distance whilst motion-based features will remain discernible. Thus to achieve full automation, existing systems will have

to combine both appearance and motion features. The aim of this thesis is to contribute to this problem by developing computer vision techniques which combine appearance and motion features to robustly classify species, whilst in flight. It is believed that once this is achieved, with additional development, it will be able to support the surveying of species and their behaviour studies.

The first focus of this research was to refine appearance features previously used in other related works for use in automatic classification of species in flight. The bird appearances were described using a group of seven proposed appearance features, which have not previously been used for bird species classification. The proposed features improved the classification rate when compared to state-of-the-art systems that were based on appearance features alone (colour features).

The second step was to extract motion features from videos of birds in flight, which were used for automatic classification. The motion of birds was described using a group of six features, which have not previously been used for bird species classification. The proposed motion features, when combined with the appearance features improved classification rates compared with only appearance or motion features.

The classification rates were further improved using feature selection techniques. There was an increase of between 2-6% of correct classification rates across all classifiers, which may be attributable directly to the use of motion features. The only motion features selected are the wing beat frequency and vicinity features irrespective of the method used. This shows how important these groups of features were to species classification. Further analysis also revealed specific improvements in identifying species with similar visual appearance and that using the optimal motion features improve classification accuracy significantly.

We attempt a further improvement in classification accuracy, using majority voting. This was used to aggregate classification results across a set of video sub-sequences, which improved classification rates considerably. The results using the combined features with majority voting outperform those without majority voting by 3% and 6% on the seven species and thirteen classes dataset respectively.

Finally, a video dataset against which future work can be benchmarked has been collated. This data set enables the evaluation of work against a set of 13 species, enabling effective evaluation of automated species identification to date and a benchmark for further work in this area of research. The key contribution of this research is that a species classification system was developed, which combines motion and appearance features and evaluated it against existing appearance-only-based methods. This is not only the first work to combine features in this way but also the first to apply a voting technique to improve classification performance across an entire video sequence.

Acknowledgements

First of all, I am grateful to the Almighty God for His guidance and protection throughout this research.

I am very grateful to my supervisors Dr. Patrick Dickinson and Dr. Wenting Duan, who have provided inspiring supervision, continued guidance, unconditional support and encouragement throughout this research. I have greatly appreciated their immense wisdom. I am also grateful to Dr. John Murray for his support and advice when I started this research. This thesis would not be possible without their trust. I have greatly appreciated my colleague, Dr. Kofi Appiah and would like to convey my special thanks to him for his valuable advice on my research.

I am also grateful to my examiners Prof. Bob Fisher and Dr. Bashir Al-Diri for reviewing this work. Their comments and suggestions immensely helped to improve this thesis.

Some special people are not mentioned yet, because they deserve their own part: Edward Shaw, Peter Cowling, Magi and Peter Rayner. I praise the enormous amount of support, ecological input, project guidance and advice provided by Dr. Edward Shaw throughout the past two years. A special acknowledgement goes Magi and Peter Rayner for their valuable time and efforts in editing this thesis.

A very special thanks to Peter Cowling of the Lincolnshire Bat Group. Your initial bat data provided a jump start for this thesis. Sincere thanks to The National Parrot Sanctuary, Lincolnshire, UK, who have enthusiastically supported this work by assisting with the collection of video data of several species used.

I gratefully acknowledge the partial funding for this study provided by the University

of Lincoln. Last but not least, I would like to thank my family (Ivy-Carol, Chaliwen, Wentame and Charles), without their endless support, patience, love and encouragement this thesis simply would not have been possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Bird species are recognised as important biodiversity indicators (Gregory, 2006; Harrison et al., 2014; Buckland et al., 2012): they are responsive to changes in sensitive ecosystems, whilst population-level changes in behaviour are both visible and quantifiable. They can be monitored by ecologists to determine factors causing population fluctuation and to help conserve and manage threatened and endangered species. For example, a study by Mikusiński et al. (2001) has shown that where Woodpecker (*Picidae*) species are present there are also other species present. The research therefore suggested that surveying the number of Woodpeckers can serve a proxy for avian diversity if the overall species diversity in the forest is unknown. Also, a study of butterfly and bird species (Blair, 1999) showed that occurrence of the two were correlated, that is, the number of birds in a particular area could also indicate the likely presence of butterflies. Data about bird populations is therefore an important tool for ecologists in a wide range of environments and contexts, including farmland use, marine settings, and migration behaviour (Hammers et al., 2014; Johnston et al., 2014; Goodenough et al., 2014).

There are a number of regular surveys conducted in the UK which provide data (Baillie et al., 2014; Robinson et al., 2015) used by the government and other organisations to track trends in bird breeding and migration data. This helps to measure progress towards international targets set by the Convention on Biological Diversity in 2010, (Gregory et al., 2015). The monitoring of bird species is also a statutory requirement governed by the Wildlife and Countryside Act 1981, the European Union Birds Directive, and the

Ramsar, Bonn and Berne Conventions (Gregory et al., 2000).

These surveys have previously included the Common Birds Census and Nest Record Scheme, which were organised annually by the Joint Nature Conservation Committee (JNCC) and attracted more than 2500 volunteers. These were succeeded in 2001 by the Breeding Bird Survey which is organised jointly by the British Trust for Ornothology (BTO), JNCC and Royal Society for the Protection of Birds (RSPB). This current scheme has over 2,800 volunteers contributing to it anually. The Wetland Bird Survey is also carried out by volunteers at more than 2,200 wetland sites, at monthly intervals, and provides additional information on wintering population trends in species of water birds including ducks, geese, swans, waders, grebes, rails and cormorants (Robinson et al., 2015). Finally the Big Garden Birdwatch attracts well over 312,000 volunteers annually, spotting over 6,295,000 individual birds throughout the year.

All of these surveys provide data, collected by volunteer and staff workers, using manual techniques. However, manual techniques are labour intensive and error-prone. For example in 2015, the BTO reported that *"The first Breeding Bird Survey (BBS) volunteers, who surveyed their squares 20 years ago, would have counted twice as many migratory cuckoos and whinchats as they do today."* (Hayhow et al., 2014). This was due to the fact that flocks often contain several species and could be very difficult to count by volunteers manually. Another important issue relates to protected bird species (including Barn Owl (*Tyto alba*), Kingfisher (*Alcedines*) and Little Tern (*Sternula albifrons*)), which by law require a handling licence: this places further constraints on monitoring activities and techniques, which are particularly important in these cases as protected species are already in decline.

Every five years, the health of the bird population found in the UK is reviewed by the BTO, based on data gathered about bird populations from these surveys: this has been a key function of the BTO since its formation in 1933 (Robinson et al., 2015). Species are categorised using three lists (Red, Amber and Green), which indicate the strength of their populations, nationally. The last review report (Eaton et al., 2015) included reviews of a total of 244 species, with 20 moving onto the red list and only three leaving it.

Bird species are therefore established as important indicators of national (and global) biodiversity and enviornmental conservation, and a considerable amount of effort is spent in gathering processing data about their populations. However, data collection techniques remain largely dependent on traditional manual observation methods, which is a limiting factor. The work presented in this thesis aims to contribute new technological methods of addressing these limitations.

1.1 Current Data Collection Techniques

A number of more technologically advanced techniques are sometimes used to survey bird species, which includes bioacoustics counting (Lopes et al., 2011; Briggs et al., 2012; Tan et al., 2015; Silla et al., 2013; Joly et al., 2014; Evans and Rosenberg, 2000), Radar for detecting migratory flocks (Gregory et al., 2004), trapping using mist nets (Gregory et al., 2004), and thermal imagery (Matzner et al., 2015; Betke et al., 2008; Ammerman et al., 2009; Hristov et al., 2010; Cullinan et al., 2015).

Bioacoustic methods use specialised equipment such as infrasound, detectors of ultrasound or laser vibrometer to capture and model the sounds of different bird species. The captured sounds are then processed using signal processing, data mining and machine learning techniques for classification. The disadvantages of these technique are that:

- specialised skills are required to operate expensive equipment.
- some bird species emit little or no sound, and thus monitoring them using these techniques may become impractical.
- Signals are often noisy and difficult to differentiate. For example mixed-species foraging flock, which may consist of Blue, Great, Coal and Long-tailed Tits, Nuthatches, Goldcrests, Tree Creepers and Lesser Spotted Woodpecker.

Radar is another technology which has been used to estimate size of flocks, their flight direction, speed and wing beat frequencies (Gregory et al., 2004). It makes use of expensive equipment, which is unavailable to most researchers. Bird detection with radar

is also challenging due to noise from electronic components of the radar system (Cullinan et al., 2015), and, in any case, skilled personnel are required to operate these systems.

Thermal imaging technologies have also been used to survey birds (Matzner et al., 2015; Cullinan et al., 2015). The use of thermal imaging to detect species may be limited by excessive humidity which affects the clarity of the image, distance from the camera, field of view and physical obstruction (Cullinan et al., 2015). Even though thermal cameras have generally lower resolution than optical, they are suitable for both day and night recordings but are again very expensive. In addition, many features used for species classification are not apparent using thermal imaging.

1.2 Using Computer Vision to Monitor Bird Species

The use of computer vision techniques to automatically identify, classify and monitor bird species in the field is an emerging area of research. Currently, most work in this area focuses on the identification of species from a single image (Marini et al., 2013; Wah et al., 2011a; Duan et al., 2012; Berg and Belhumeur, 2013; Huang et al., 2013; Branson et al., 2014; Wah et al., 2011b; Berg et al., 2014), using features extracted from that image which represent the appearance of the bird, such as colour and shape. Datasets for these works are based on high-quality and detailed images or stationary birds. However, in the field, images taken by biologists, ecologists or camera traps may not be of comparable quality, placing some practical limitations on such approaches in the field.

These type of approaches can be further subdivided into those that make use of information about the physical structure of individual birds (which is refer to as "part-based") and those which do not. Non part-based methods use colour and shape features of the bird as a whole to classify its species (Marini et al., 2013; Wah et al., 2011a,b). For example the work in Marini et al. (2013) used colour features extracted from the bird to build a Support Vector Machine (SVM) species classifier. Most of these works have been used to classify between relatively small numbers of species but struggle to maintain peformance as the number of species increases. Marini et al. show that in classifying species using colour features alone on the caltech-ucsd birds-200-2011 Dataset, accuracy reduces

from approximately 85% when selecting between 2 species to 20% when differentiating between 17 species, and just 7% accuracy with 200 species.

Part-based methods use features associated with the various parts of the bird, based on colour and/or shape (Wah et al., 2011a; Duan et al., 2012; Berg and Belhumeur, 2013; Huang et al., 2013; Branson et al., 2014; Wah et al., 2011b; Berg et al., 2014). This approach uses prior-knowledge of the birds appearance and can help differentiate between species with high visual correlation. For Example Collared and Turtle Dove (*Streptopelia decaocto* and *Streptopelia turtur*) are species with distinguishing features around the neck and the eyes: Turtle Doves have a black and white striped neck patch, and a bold red eye ring, which are not visible on Collared Doves. Other notable examples are warblers and vireos, which have distictive wing markings, compared with flycatchers and sparrows which have no such markings (Podulka et al., 2004). In such species these methods have achieved good classification accuracy but do require some manual inputs (so are not fully automated) and good-quality images in which all parts are present. Typically the manual inputs are annotations which identify the bird's parts prior to feature extraction, and this is time consuming and labour intensive, placing some practical limits on the amount of data which can be processed. Gavves et al. (2013) introduced some improvments in the form of automatic parts identification, however their system still requires some manual input being dependent on the grab cut method (Rother et al., 2004a). Techniques based on parts are also less applicable to classificastion of flying birds, which have less well-defined object shapes and in which specific parts (e.g. feet) are typicaly obscured from view.

Flight patterns are known to vary across different species of birds (Briggs et al., 2012). For example woodpeckers generally fly in patterns of moderate rises and falls whereas finches exhibit a steep, roller-coaster flight (Podulka et al., 2004). Flight patterns can sometimes help bird watchers distinguish species with similar colour and shape, for example Common Raven (*Corvus corax*) and the American Crow (*Corvus brachyrhynchos*) (Kilham, 1990). Characteristic flight patterns are particularly helpful to human observers and aid the identification of bird species in flight, especially at distances where

colour tends to attenuate and shape features are too small to distinguish.

As mentioned, existing computer vision based approaches to automated species classification focus on single image inputs and appearance features. The automated classification of birds in flight has remained largely un-studied, due to problems with image quality and extraction of appearance features. However, the use of video instead of single images presents the opportunity to use motion features. Very little existing work has considered this, and the primary objective of this research is to combine motion features with appearance features to provide robust automated classification of birds in flight: a technology which would be immensely useful for deployment in the field. The only existing comparable works are those presented by Duberstein et al. (2012) which address only broad categorisation (bats, swallows, terns and gulls) and use thermal sensors and also work by Matzner et al. (2015). However, both these works are very limited in that the number of species/categories, the data sets are small, and the species used in the experiment have obviously very different flight patterns. The objective is to build on these preliminaty studies to develop a method which can be used robustly with larger numbers of species and give similar or better results as methods which use single high-quality images.

This thesis presents work which combines motion and an extended set of appearance features to automatically classify birds in flight from video streams. The research shows that robust results across larger numbers of species is possible and that motion and appearance features can be combined effectively to achieve this, providing a robust platform for deployment in the field.

1.3 Aims and Objectives

As mentioned in the last section, the aim of the work presented in this thesis is to develop computer vision techniques which combine appearance and motion features to robustly classify a bird's species, whilst in flight. More specifically:

Objectives

- To evaluate appearance features used in previous related works and refine them for

use in automatic classification of species in flight.

- To identify relevant motion features which can be extracted from video of birds in flight and used to classify them automatically.
- To effectively combine appearance and motion features and evaluate whether this combination gives better results than appearance or motion features alone.
- To eliminate redundant and irrelevant features from this aggregated feature set, to determine if recognition can be further improved.
- To establish a video data set of birds in flight for use by the research community.
- To investigate the further refinement of this technique, using majority voting to aggregate classification results across a set of video sub-sequences.

1.4 Challenges

There are a number of important challenges inherent in these objectives. Specifically:

- Identification of a suitable set of appearance features: Most existing works have used appearance-based features in conjunction with single images of static birds. This research develops an anlogous set of appearance features for use with birds while in flight. Whilst existing works are a useful starting point, such as those used by Marini et al. (2013), there is no comparable work for in-flight birds.
- Identification of suitable motion features: Even though these have been explored for the classification of fish trajectory (Beyan, 2015; Wang et al., 2015) and human motion identification (Cutler and Davis, 2000; Ayyildiz and Conrad, 2011), there is no significant work which considers bird species. The identification of a suitbale set of motion features is an open question and there represents a significant challenge.
- Development of a framework to combine motion and appearance features: Motion features are extracted from a sequence of image frames while appearance features are extracted from individual images. Determining an effective way to combine them for classification is challenging in the absence of comparable exisitng work.

- Identification of which features (motion and appearance) are most effective: This research starts by combining all potentially useful appearnce and motion features, resulting in a large feature set. Determining how best to reduce and optimise this feature set presented a significant challenge.

- Segmentation of birds in flight: motion features are extracted by fitting a bounded box to a detected bird in flight, based on the bird's silhouette. Sudden light changes and low video resolution directly affect the motion feature extracted to form the trajectories.

- Collection of an effective dataset: There is no pre-existing standard data set of birds in flight, on which to test the algorithms and methods developed in this research. It has been determined that high frame-rate video would be most useful for extracting motion features and consequently needed to developed data set for this research. It is challenging to find suitable locations to film different species, and produce an effective data set with which to work.

1.5 Original Contributions

The work presented in this thesis makes the following original contributions to this field of research:

1. The development of algorithms for the measurement of wing beat frequencies using Fast Fourier Transform (FFT) which have been evaluated initially on bat species and then on bird species. The techniques used in this research were evaluated with the state-of-the-art, Cutler and Davis (2000) and the results show that they performed significantly better.

2. The development of a new set of appearance features which show improved correct classification rate over the state-of-the-art proposed by Marini et al. (2013) on the dataset of birds in flight. The appearance features outperform Marini *et. al.* by between 6% and 14% across various datasets.

3. The development of a species classification scheme which combines motion and appearance features and evaluation of it against existing appearance-only-based methods. This is the first work to combine features in this way. A correct classification rate of 85% was achieved which outperforms the exisiting state of art method (Marini et al. (2013)).

4. Feature reduction techniques were further used to improve classification rate on the combined features and demonstrated the relative importance of motion features within this reduced set. Two feature reduction techniques were evaluated against the full feature set, which have shown approximately a 4% increase in performance.

5. A voting technique was used to improve classification performance across an entire video sequence. This has increased classification rate to between 89% and 98%.

6. A video data set was collated against which future work can be benchmarked. This data set enables the evaluation of the work against a set of 13 species, enabling effective evaluation of automated species identification to date and a benchmark for further work in the field of research.

1.6 Publications

Some of the work described in this thesis has been previously presented in journals and conference papers. Below is a complete list of publications arising from this PhD study so far.

Journal Paper

1. Atanbori, J., Duan, W., Murray, J., Appiah, K., & Dickinson, P. (2016). *Automatic classification of flying bird species using computer vision techniques*. *Pattern Recognition Letters*, ISSN: 0167-8655, DOI: http://dx.doi.org/10.1016/j.patrec.2015.08.015. Volume 81, Pages 53-62.

Conference Papers

1. Atanbori J, Cowling P, Murray J, Colston B, Eady P, Hughes D, Nixon I & Dickinson P. (2013, January). *Analysis of bat wing beat frequency using Fourier transform. In Computer Analysis of Images and Patterns* (pp. 370-377). Springer Berlin Heidelberg.

2. John Atanbori, Wenting Duan, John Murray, Kofi Appiah and Patrick Dickinson. *A Computer Vision Approach to Classification of Birds in Flight from Video Sequences*. In T. Amaral, S. Matthews, T. Plotz, S. McKenna, and R. Fisher, editors, *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB)*, pages 3.1-3.9. BMVA Press, September 2015.

1.7 Organisation of the Thesis

The remainder of this thesis is structured as follows:

Chapter 2 gives a comprehensive overview and comparison of existing works in the area of automated systems for species classification. This not only includes bird species but also other animal species that have been studied using computer vision techniques. An evaluation of literature based on species identification and classification using bioacoustics, appearance, and motion features was also presented. Finally, literature on other relevant work in the area of feature selection, evaluation using imbalanced datasets, machine learning and classification techniques were presented.

Chapter 3 presents initial work on extraction of motion features. For this first study video data of bats collected from bat roosts rather than birds were initially used. Bats were initially used for this study partly because the data was readily available but also because bat motion is of a higher frequency and so is more challenging for analysis. The chapter also presents a discussion of low-level image processing techniques for dealing with low-light video data of bats and bird species. A number of techniques were introduced for the analysis of bat wing beat frequency and the results from this methods were

evaluated with state-of-the-art method by Cutler and Davis (2000).

Chapter 4 presents work on the classification of bird species in flight, using only appearance features. In this chapter the dataset used in all evaluations were detailed and statistical features that were used to represent the feature sets used in this research were introduced. The work presented here uses a rich set of appearance features with standard classifiers (Naive Bayes, Support Vector Machine, Random Decision Trees and Random Forest) to classify species. Three datasets (The seven species (Dataset #1), The thirteen classes (Dataset #2) and caltech-ucsd birds-200-2011 Dataset) were used to evaluate the work presented in this chapter. The results from the appearance features presented in this chapter compares favourably with exiting state-of-the-art image-based classifiers used in the work by Marini et al. (2013). This out performs the state-of-the-art (Marini et al., 2013) on all 3 datasets: 9% on caltech-ucsd birds-200-2011, 6% on Dataset #1 and 9% on Dataset #2.

Chapter 5 presents the work on fusing appearance and motion features. A rich set of motion features were used with the appearance features from chapter 4 with standard classifiers (Naive Bayes, Support Vector Machine, Random Decision Trees and Random Forest) to classify bird species. Dataset #2 was used to evaluate the work presented in this chapter. Using only motion features for classification, a classification accuracy of 38% was achieved. Fusing these features with the appearance features in this chapter, a classification accuracy of 85% was achieved. The result was compared with that of using only appearance features, and an initial slight reduction in classification accuracy by approximately 1% was reported, thus, motivating the work in the following chapter.

In Chapter 6 the results of work on feature selection was presented. Combining appearance and motion features in Chapter 5 resulted in a large set of features, which slightly degraded performance. To improve classification, feature selection techniques were applied to remove redundant features. The most important features were identified using

correlation and classifier based feature selection techniques. These were used to classify species using the Dataset #2 with the four standard classifiers (Naive Bayes, Support Vector Machine, Random Decision Trees and Random Forest) and a correct classification accuracy of 89% was attained, which is approximately 4% better than that in Chapter 5. A further experiment was performed to determine the contribution of the selected motion features to overall performance. The conclusion was that the motion features used together with appearance improves classification by approximately 4% across all four standard classifiers.

In Chapter 7, further improvement in the correct classification rates was attempted. The work in chapters 4 - 6 present results based on classification using single frames and subsets of video. This was extended to combine the results of several frames from a sequence using majority voting. It has been established that combining the outputs of several classifications result in a better overall accuracy (Bhattacharya and Chaudhuri, 2003). This is because different classifications can capture different aspects of the input data, while one alone will not usually represent all. The work presented here uses the results in Chapter 6 with majority voting and the four standard classifiers (Naive Bayes, Support Vector Machine, Random Decision Trees and Random Forest) to classify bird species. The datasets used included that with the seven species and the extended thirteen classes dataset.

Finally, Chapter 8 discusses the results and the contributions and limitations of this thesis and outlines future works.

Chapter 2

Literature Review

The previous chapter mentioned the importance of monitoring bird species. In particular, it was discovered that ecologists monitor them to determine the factors causing population fluctuation and to help in conserving and managing threatened and endangered species. The various surveys used in counting bird species including data collection techniques were succinctly reviewed. It was established that a small but growing number of researchers have studied the use of computer vision for monitoring species, particularly for counting bat species.

This chapter evaluates reports of studies found in literature that are related to monitoring and classification of species. In particular, it focuses on reviewing birds, fish and bats works: techniques used for these species are often similar, and motion features which this research seeks to investigate for classification of birds have been used for both bats and fish. First techniques which perform classification using single images were explored. This was done by reviewing them separately as those that are used for classification of bird species and those for other species, specifically, bats and fish.

The work based on bird species was then splitted into two sections: part-free and part-based models. Feature selection and machine learning techniques, relevant to the classification of these species were separately reviewed. Finally, a brief overview of video classification system using computer vision techniques was presented. This chapter is structured into the following sections:

- Section 2.1 reviews bio-acoustics for classifying bird species

- The classification of species using computer vision techniques were reviewed in Section 2.2, by first looking at literature related to bat and fish species classification and monitoring, and finally, those related to bird species.
- Section 2.4 and 2.5, review feature selection and reduction methods and imbalanced datasets techniques respectively.
- Finally, machine learning algorithms for classification were reviewed in Section 2.7 and an overview of the video classification system in Section 2.8

2.1 Classification of Bird Species using Bio-acoustics

A number of existing attempts to automate the identification of birds have used audio rather than visual signals, such as Briggs et al. (2012); Neal et al. (2011); Bardeli et al. (2010). In particular, Briggs et al. classified 413 birds songs, each of 30 seconds, using FFT and Nearest Neighbour (NN) classifiers. This achieved a remarkable result (92%), which was due to the Fast Fourier Transforms (FFT) used, filtering out most of the noise in the signal. Neal et al. also proposed a supervised time-frequency audio segmentation to extract syllables of bird calls. They then applied Random Forest (RF) to classify 625 birds songs and achieved a correct classification rate of 83.6%. Even though the number of species used in Neal et al. was larger than that of Briggs et al., the improvement in classification was attributed to the state of the art Random Forest classifier used. Random Forest has a mechanism for boosting results by splitting the dataset into random trees and applying majority voting, which helps improve classification rates.

All the above research have used FFT to improve the quality of the signals, as audio recordings taken in the field are usually buried in noise. It has been demonstrated by Lopes et al. (2011) that when the number of species is increased the classification rate reduces, even with FFT. They performed experiments using 3, 5, 8, 12 and 20 species and by first applying FFT to the signals to reduce the noise and then applying a sound ruler software, with Naive Bayes (NB) classifier to classify species by vocals. Lopes et al. showed that by randomly using 3 species and averaging the results, the classification rate was 61.5% whilst this was 25.4% for 20 species. Other work using vocals did not attempt

classification. For example, Bardeli et al. (2010) developed a real-time bird-counting system using their vocals.

The use of audio signals has some attractive features; species typically have distinctive calls and no line of sight is necessary to detect audio. However, there are also a significant disadvantages. Firstly, audio signals are sometimes buried in noise and it becomes difficult if not impossible to remove the noise. Secondly, audio signals are sparse (an individual may emit no audio at all for extended periods) and it is unrealistic to differentiate some individuals in this way (e.g. for species-level counting).

Due to the limitations of using the audio signals for classification, one work (Marini et al., 2015) attempted to improve species classification rates by combining appearance features from the CUB200-2011 dataset (Wah et al., 2011b) with audio signals from the Xeno-Canto dataset. This work achieved an improved correct classification rate of between 1.2% to 15.7% for species when compared to using appearance features alone. Even though this has improved classification rates, it was for a limited number of species and recognition rates were not very significant compared to methods that uses appearance features alone (Berg and Belhumeur, 2013; Branson et al., 2014; Berg et al., 2014). The limited number of species used was due to the fact that the CUB200-2011 dataset did not have enough corresponding audio counterparts in the Xeno-Canto dataset.

2.2 Application of Computer Vision Techniques to Species Classification

Whilst several species recognition systems have used audio, more recently, recognition systems have focused on the application of computer vision techniques. One work that used both computer vision and bio-acoustic processing is the work in Marini et al. (2015) but the results in this case was no different to those that use just computer vision techniques (Berg and Belhumeur, 2013; Branson et al., 2014; Berg et al., 2014). Therefore, since this thesis focuses on using only computer vision techniques in the classification of species, the following sections of this chapter have been dedicated to the review of computer vision techniques.

2.2.1 Studies related to Other Species

Whilst computer vision applications to bird species represent the most relevant existing works, there have been a number of other studies which have been applied to bats and which address similar problems and challenges. Unlike birds, bats are predominantly nocturnal and applications of computer vision techniques need to be effective in low-light conditions. In particular, colour features are generally undetectable, and existing approaches have therefore looked more closely at the extraction of motion features.

Existing works, such as Cullinan et al. (2015); Matzner et al. (2015); Hristov et al. (2010); Betke et al. (2008); Lazarevic et al. (2008) have been motivated mainly by automated censusing of large populations, as well as classification of individuals. A small number of these works, such as that by Betke et al. (2008), have attempted to estimate wing beat frequencies of individual bats. Betke et al. used thermal imaging together with pose templates to form a signal which was then analysed using Fast Fourier Transform (FFT) techniques to recover a frequency estimate of signals buried in noise. Hristov et al. also used a thermal camera with computer vision algorithms to detect and track species. Species that were successfully tracked beyond a set threshold value were counted automatically to estimate the total number of emerging bats from a cave.

Like Betke et al. the works by Cullinan et al., Matzner et al., and Lazarevic et al. also use thermal imaging, but include some attempts at classification, including broad species of birds. Cullinan et al. attempted to identify four categories (bat, gull, tern, swallow) using flight tracks only. They reported 82% correct classification; however, this represents only a small number of categories, does not consider fine-grained differentiations and is highly contextualised to offshore wind-farms. Similarly, Lazarevic et al. were also motivated by studies around wind farms but looked at only coarse classification (bat or bird). Some disadvantages common to all these methods are that thermal imaging for detecting species may be limited by excessive humidity which affects the clarity of the image, distance from the camera's field of view (Cullinan et al., 2015). Recordings from these cameras have generally lower resolution compared to optical counterparts, which makes them more suitable for both day and night recordings, but may again be very expensive.

Finally, many features used for species classification are not apparent using thermal imaging.

Unlike bats, which are typically filmed in low-light, the automated monitoring and classification of other species is able to make wider use of appearance features (such as colour). Some recent attempts to automate the classification of fish species have used both appearance and motion features (though not combined). For example, Lee et al. (2003) used shape contour features to discriminate between nine different fish species achieved an overall true positive rate of between 13% and 80%. Spampinato et al. (2010) also developed a fish classification system which used texture and shape features, achieving 92% overall true classification rate; the combination of different appearance features was effective in this case. Another appearance-based technique by Rodrigues et al. (2010) used Scale-Invariant Feature Transform (SIFT) and Principal Component Analysis (PCA) to select effective linear combinations of features; the results of their evaluations were similar to those by Lee et al. and Spampinato et al., achieving an overall true positive rate of 92% across six species. More recently, trajectories (motion) features from video streams have been used to model fish behaviour, and detect outliers; for example, the work of Spampinato et al. (2014), and Beyan (2015).

Fish studies using computer vision techniques have gained some popularity recently. Several datasets have been introduced to support their study (Jager et al., 2015; Fisher et al.; Joly et al., 2015). Jager et al. (2015) introduced an imbalanced dataset for classification of Croatian fish species. This is a fine-grained fish dataset, which contained 12 species of fish, made up of 794 images. The work in Jager et al. then combined convolutional Neural Networks (CNN) and SVM to perform a baseline experiment achieving a correct classification rate of 66.78%. The Fish4Knowledge dataset (Fisher et al.) was introduced for the classification and study fish behaviour. This large dataset was originally collated for the study of fish behaviour studies but has a great potential for classification, with a potentially large number of species.

A prolific number of datasets for the study of species is an indication of how important the application of computer vision to animal study is becoming. A butterfly dataset

(Anwar et al., 2015b,a), with some initial investigations into the classification of species has also been developed. Anwar et al., used a bag of visual words to classify 30 species by including spatial information of extracted images. Other butterfly datasets include Lazebnik et al. (2004) which is made up of 7 species and the Leeds butterfly dataset Wang et al. (2009) made up of 10 species.

2.2.2 Bird Species Classification using Appearance Features

Whilst Cullinan et al. (2015), and Matzner et al. (2015) mentioned above have attempted to use motion features to differentiate between small numbers of species, all other existing works concerned with the automated classification of birds use appearance features derived from a single image of an individual bird. These type of approaches can be further subdivided into those that make use of information about the physical structure of individual birds (referred to as *part-based*), and those which do not. Non-part-based methods use colour and shape features of the entire bird, without considering the relative position, to determine its species (Marini et al., 2013; Wah et al., 2011a,b). For example, the work by Marini et al. uses colour features extracted from the bird with a Support Vector Machine (SVM) classifier. Again most of these works have been used to differentiate between relatively small numbers of species and struggle to maintain performance as the number of species increase. Marini et al. showed that when using colour features alone on the Caltech-ucsd birds-200-2011 Dataset, accuracy reduces from approximately 85% when selecting between 2 species to 20% when differentiating between 17 species and just 7% accuracy with 200 species. SIFT and colour features (Wah et al., 2011b) work well for the classification of bird species but again, is only tested for a small number of classes. The work presented in Welinder et al. (2010), classified bird species using size and colour histogram with a bin size of 10 but the classification rate was also low, which was attributed to the fine-grained nature of the dataset.

Part-based methods use features which are associated with specific parts of the bird's body and use colour and/or shape features (Wah et al., 2011a; Duan et al., 2012; Berg and Belhumeur, 2013; Huang et al., 2013; Branson et al., 2014; Wah et al., 2011b; Berg et al., 2014). This general approach can help differentiate between species with high

visual correlation. For Example Collared and Turtle Doves (*Streptopelia decaocto* and *Streptopelia turtur* respectively) are species with distinguishing features around the neck and the eyes but Turtle Doves exhibit a black and white striped neck patch and a bold red eye ring, which is not visible on Collared Doves. In such species, these methods have achieved better classification accuracy but almost all require some manual input (so are not fully automated) and also require good-quality images in which body parts are present and identifiable. Typically, the manual inputs are annotations which identify the bird's parts prior to feature extraction and this is time-consuming and labour intensive, placing some practical limits on the amount of data which can be processed. For example, Branson et al. proposed a *human in the loop* approach, which is predicated on the idea that a human and computerised system working together classify more efficiently than either alone. The human operator annotates the parts and answers multiple choice questions and the algorithm uses this information to assist with the classification: they were able to achieve a true positive rate of 93%. Berg et al. developed an online application called Birdsnap, which can be used to classify various US bird species; this also requires some manual input to annotate the bird's parts prior to segmentation and classification.

Krause et al. (2015), and (Gavves et al., 2013, 2015) both developed annotation-free parts-based methods, which automatically identified body parts using co-segmentation and alignment. Their results compared favourably to other states of the art methods: Krause et al. used figure/ground segmentation to determine pose and localise the bird's parts, while Gavves et al. fitted ellipse to the segmented object to align and then determine sub-parts. Results based on the CUB-2011 dataset show that the true positive rates were 62%, 82% and 67% for the methods in Gavves et al. (2013); Krause et al. (2015); Gavves et al. (2015) respectively. The drawbacks of these methods are that they make use of the Grab-Cut segmentation method Rother et al. (2004b) so still require some manual inputs: decomposing soft-bodied objects with arbitrary poses, remain a challenging problem in computer vision. Another annotation free method was proposed by Zhang et al. (2015), this detects parts using CNN feature representations. The main difference between this and Krause et al. is the method by which the parts are selected. Whereas Krause et al.

align the co-segmented objects before labelling parts, Zhang et al. (2015) uses CNN feature representations to detect parts automatically. The correct classification rate reported by Zhang et al. was (75%) based on the CUB-2011 Dataset and was not a significant improvement on the other methods reported.

The objective of this research is to develop a system which is capable of identifying birds in flight, which can be deployed in the field. Almost all of the part-based methods mentioned above are dependent on manual annotations and whilst many have been successful, this constraint makes them inherently unsuitable for wide-scale deployment in the field. Further more, they all also require high-quality images. The datasets used have been highly-detailed, high-resolution images which exceed the quality that would be captured automatically from flying birds in real-world settings. In addition, it is asserted that birds furthest from the camera would be less easily classified using appearance features alone (for example, colour features are attenuated). This motivates our approach of combining colour and motion features. Of the other approaches mentioned, it was consider that Marini et al. is most appropriate for this problem domain, being fully automated, non-parts based (so more likely to be robust to reduced image quality) and reporting relatively good results. This method have therefore used as a benchmark for the work presented in this thesis.

2.2.3 Bird Species Classification using Motion Features

All the methods in Section 2.2.2 use single images and appearance-based models for classification; however, bird species also exhibit distinguishing behaviours (flying, moving, poses, etc) which could also be used to help robust automated identification. This is particularly relevant to the identification of birds in flight, especially at a distance where appearance-based features such as colour tend to attenuate, whilst motion-features remain discernible. In ecological studies, conservation and biology, motion-based vision techniques have been applied to tracking flying species and analysing their kinematics, behaviours, and flight trajectories (Breslav et al., 2012). The analysis of bats motion using computer vision techniques have widely been studied by researchers for application in ecological, conservation and biology. Bats are nocturnal creatures and are active at night

and hence most studies have used thermal cameras for the analyses of their motion. Currently, research concern with bats is mostly focusing on wing beats modelling (Breslav et al., 2012) and counting bats as they emerge from caves (Hristov et al., 2010; Betke et al., 2008).

The most significant relevant studies using motion features with bird species include Duberstein et al. (2012), which explores wing beat frequencies of bird and bats species but does not use these features for classification. Duberstein et al. (2012) like Cullinan et al. (2015), and Matzner et al. (2015) mentioned above have attempted to use motion features to differentiate between small numbers of species. Duberstein et al. (2012) explores the wing beat frequencies and flight patterns of bird and bat species by extracting descriptive statistics from flight patterns of four species of birds and bats. The statistics (which included the minimum, maximum, mean, standard deviation, quartiles and interquartile range) were used to cluster species, thus making their classifier perform faster as the number of features were small. This work was redeveloped by Cullinan et al. for broad classification of species of bats, swallows, terns and gulls. However, they only used 48 tracks from a 5 minute video from a thermal camera and these cameras may sometimes be expensive and difficult to deploy in the field. The work in Matzner et al. (2015) increased the dataset in Cullinan et al. (2015) whilst applying the same techniques in Duberstein et al. and Cullinan et al., to classify species into categories such as bats, swallows, terns and gulls.

Hitherto, works concerned with flight pattern have been carried out using recordings from tracking radar (Liechti and Bruderer, 2002; Bruderer et al., 2010; Zaugg et al., 2008) during the migration seasons of bird species while others have been carried out using thermal camera (Duberstein et al., 2012; Cullinan et al., 2015; Matzner et al., 2015). Most of these studies that use thermal cameras were recorded during the evening when species are migrating, and are also very suitable for nocturnal species like bats. Work by Huang et al. (2013) used a graphical model with saliency to classify 9 species of birds, by extracting Scale-Invariant Feature Transform (SIFT) and colour features, which were trained using different Support Vector Machine (SVM) classifiers and achieving 73.8%

classification rate, which is comparable to the results in Cullinan et al. (2015).

Characteristics of Flying Birds

Bird species can be identified by the speed at which they flap their wings, as well as their wing beat patterns. The shape of a bird's wings determines the way it flies. Smaller-winged birds tend to fly faster to maintain the same lift as those with larger wings (Cochran et al., 2008).

Most birds fly by combining flapping, gliding and soaring. The type of flight pattern used produced by a particular bird species depends on its size, weight, wing span and shape. The smaller birds usually fly using a technique of short bursts of flapping, alternated with intervals in which wings are folded against the body. This flight pattern is known as "flap-bounding" flight (Tobalske et al., 2009). They usually abandon the conventional flap-glide pattern but still glides only if they have to decelerate during flight. Examples of birds species with this characteristic flight pattern are finches and sparrows.

Some other small birds like the swift, swallow and martin glide most of the time but also occasionally combine flapping to fly faster. Other species like finches and woodpeckers have undulating flight pattern, which describes a kind of roller-coaster style where the bird flaps its wings during the rising phase, then glides as it descends.

Budgerigars fly using flap-gliding and have the tendency to fly at only two distinct fixed speeds. They switched between a high speed and a low speed suited to safe manoeuvring in a cluttered environment (Schiffner and Srinivasan, 2016). Birds like ravens and hawks have a flap-glide or flap-soar flight characteristics, which usually consist of flapping their wings with some occasional breaks from flapping by soaring or gliding. Interestingly, cockatiels blend the traditional flap-gliding with flap-bounding.

Some birds, like Parakeet, gulls, pigeons and doves have direct flight pattern, which consist of a steady flight with rapid wing beats.

General Techniques for Analysing Motion Features

Various techniques have been used for the measurement of periodic and cyclical motions, based on metrics derived from bounding boxes (Ghaderian et al., 2011), similarity matrices (Cutler and Davis, 2000; Plotnik and Rock, 2002; Lazarevic et al., 2008), object pose (Breslav et al., 2012), motion pattern (Ren et al., 2011), and point correspondence (Laptev et al., 2005). These are mainly based on the use of autocorrelation and Fast Fourier Transforms (FFT) to estimate object motion periods. In the case of human activities, Ayyildiz and Conrad (2011) used motion moments to classify videos, by calculating the image moments and fitting a 1D time domain function on the centroids and pixel variances. These were transformed into the frequency domain using FFT and achieving a classification accuracy of between 59% to 70% for 10 home activities based on 200 videos. The disadvantage of this method is that it uses a radius based classifier and the results of the classifier are highly dependent on the radius used. Ren et al. (2011), used motion templates to perform motion pattern analysis and extract periodicity information from sports videos and to classify two sports activities (weight lifting and dumbells) into qualified and unqualified. This work achieved 93.5% and 97.7% correct classification respectively. Another technique that has been used in periodicity estimation is point correspondence and RANSAC procedure, as in Laptev et al. (2005), which was used to match image sequences over a period of time in an attempt to detect periodicity. Ghaderian et al. (2011) detected periodicity in human activities by extracting image silhouettes and fitting bounding boxes around them. The distances of the targeted silhouette in four directions (top, left, right and bottom) to the bounding box sides were obtained. For example, the horizontal distances from the edge of bounding box to the contour of the object silhouette were computed and summed to form a 1D time domain function. This work was tested using 50 periodic and 50 non-periodic videos, achieving a correct recognition rate of 97%. The techniques used were prone to errors in recognition especially when segmentation error is high. This is because bounding box metrics are based on the segmentation of the species from one frame to another. The most cited paper in motion analysis is by Cutler and Davis (2000),

which used object similarity to detect periodicity of human and dogs in videos. The object periodicity was estimated by extracting foreground images and resizing them to 9x15 pixels using Mitchell Filter Schumacher (1992). The image similarity matrices were then formed from these images using absolute correlation. In order to account for tracking error, the similarity matrices were minimised using a local search radius to form recurrence matrices. A Hanning filter was then applied to the recurrence matrices, and each column transformed to their frequency domain using FFT. The power spectra were then averaged and Equation 2.1 applied to select the ideal frequency. Chapter 3 develops algorithms for the measurement of wing beat frequencies using Fast Fourier Transform (FFT), which have been evaluated initially on bat species, and evaluated with state-of-the-art techniques by Cutler and Davis.

$$P(f_i) > u_p + k\sigma_p \tag{2.1}$$

The techniques used in Cutler and Davis (2000), have also been used in Plotnik and Rock (2002); Lazarevic et al. (2008) but predominately for estimation of periodicity in animals. Plotnik and Rock (2002) applied the techniques used in Cutler and Davis (2000) to quantify motion behaviour of jellyfish to detect their motion mode changes. The only difference compared with Cutler and Davis is that they used the normalised sum of squared differences to form their similarity matrices. Lazarevic et al. (2008), used similarity matrices based on absolute correlation to differentiate airborne targets (birds and bats). To determine periodicity they used patterns produced by the similarity matrix plots. These works were more challenging than the work by Cutler and Davis in two ways: filming bats and birds, usually results had lower resolution images than human and that they had more erratic behaviours and thus measuring periodicity in these species requires a very robust technique. Recently, Breslav et al. (2012) proposed a method of computing the wing beat of bats by comparing every shape in the input shape time signal to a prototype template shape using the shape context descriptor and the Hungarian algorithm. They then scored shape poses which were used in the estimation of the wing beats. The disadvantage of this method is that it assumed bats are flying horizontally across the field of view, which

realistically is not the case as bats flight are more erratic in nature.

Motion patterns have also been studied in fish species to mainly differentiate trajectory into either normal or abnormal, but this is for a few species. Tian et al. (2014) differentiated normal and transgenic Zebrafish using the histogram of body bending along the zebrafish body, motion displacement vectors between two consecutive frames, speed and acceleration and motion between consecutive frames in the three-dimensional space over time. This work achieved a 73.99% recognition rate based on hybrid features. The Fish4Knowledge dataset introduced by Beyan and Fisher (2013b) was used in Beyan (2015) to classify fish trajectories by extracting 776 features, which was reduced using PCA to 140 features. These were used to cluster fish tracks as normal and abnormal. Fouad et al. (2013) used SIFT and SURF features with an SVM classifier to identify fish as Tilapia or non-tilapia and achieved a classification accuracy of between 56.6% to 100%. The good results acheved by this work were due to the small number of classes involved. Recently, Spampinato et al. (2014) have used video texture analysis and SVM to study differences in fish behaviour when disruptive events such as "typhoons" happen. Whiles Wang et al. (2015) analysed behaviour of fish species using a similarity-based periodicity detection combined with the K-neighbors classifier. The method by Wang et al. was, however, not very robust to noise in the video as it uses a similarity-based approach.

2.3 Related Feature Extraction Methods

2.3.1 Appearance Features

Colour histogram features have been used widely to describe colour images by extracting the histogram from various colour channels. To reduce the feature vector dimension and to make the system run in real-time, statistical measurements can be used to describe the colour histogram (Sergyan, 2008; Huang et al., 2010). A popular way to identify birds from video is using colour. Thus we deployed colour moments and calculated statistical features for speed of training and prediction.

Related to the colour features used in this thesis is the work of Marini et al., which experiments in chapter 4 were benchmarked against. Marini et al. performed two main

experiment using two different colour spaces, the RGB and HSV colour spaces. Regardless the color space used, three-color histograms are computed, one for each channel. The histograms are converted into feature vectors by a binning process where each channel is represented by a fixed number of bins (30 for hue, 32 for both saturate and value). Experiments showed that the HSV colour space performed better than the RGB, therefore this thesis used the HSV colour space.

Two different approaches were used to represent the feature vector. In the first approach, three feature vectors are concatenated and handled by a single classification algorithm as shown in Figure 2.1, whereas in the second, each feature vector is handled by a different classifier and their outputs are combined using a fusion rule to classify bird species (Figure 2.1).

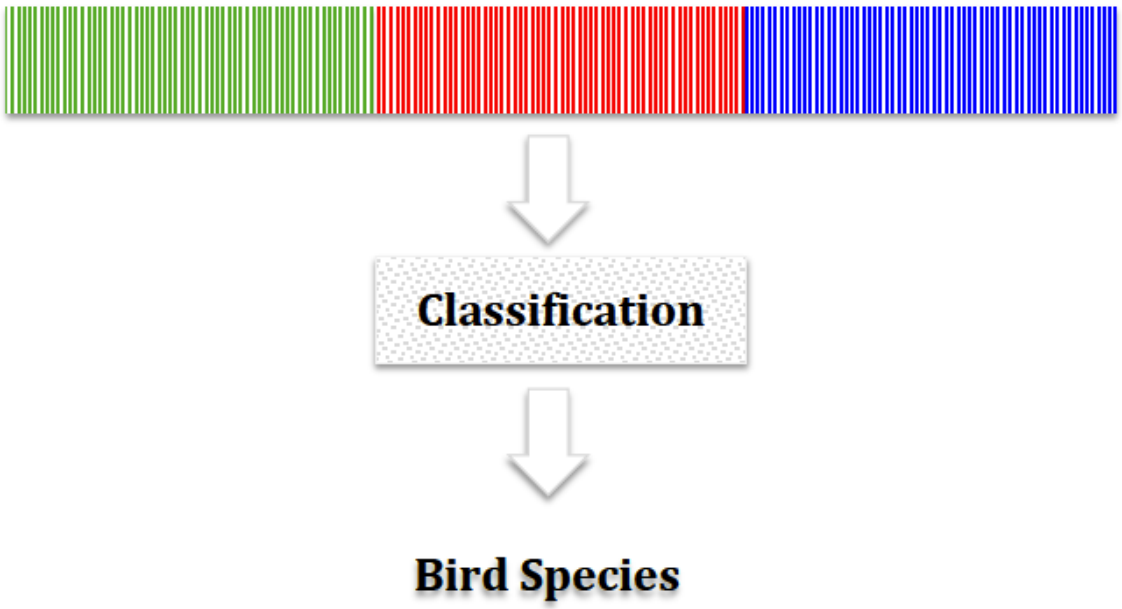


Figure 2.1: Classification scheme based on the concatenation of the three color space histogram features - Adapted from Marini et al..

Classification was performed using a support vector machine (SVM), as it more efficiently and can perform non-linear classification using a kernel trick. The SVMs classifiers was implemented with a Radial Basis Function kernel and the gamma and cost parameters optimized using grid search. The one-against-one SVM was deployed which builds $\frac{K(K1)}{2}$ binary classifiers for a K multi-class problem and class prediction is determined by majority voting by combining the outputs of all $\frac{K(K1)}{2}$ classifiers.

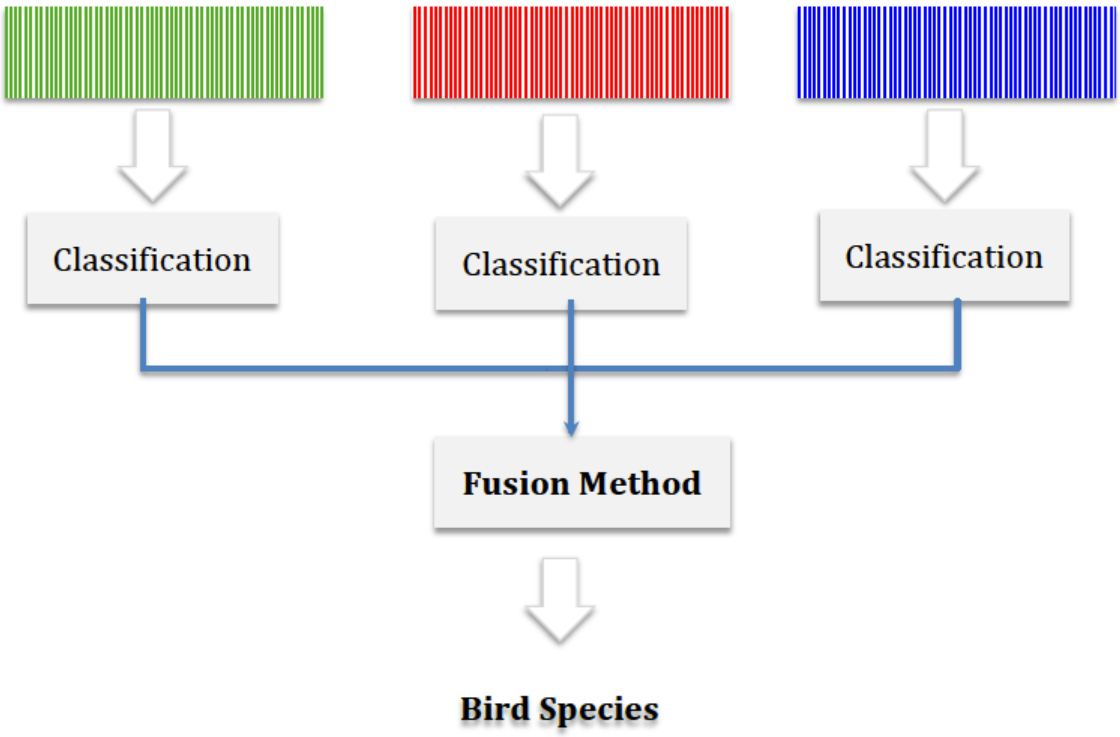


Figure 2.2: Classification scheme based on combination of classifiers outputs from the three colour space separately - Adapted from Marini et al.

Shape Features is another important feature for identifying birds species. To describe the shape of an object various image moments can be extracted from the image contours (Du et al., 2007). An image moment is a weighted average of the image pixels' intensities. We used two moments in the extraction of shape features for bird species classification, which include spatial (raw) moments and Hu moments (Fig. 4). Hu moments are invariant to some transformations, such as rotation, scaling, and translation (Martín et al., 2010) and are therefore well suited for flying bird species classification. Assuming a grayscale image with intensity values $I(x,y)$, the raw image moments M_{ij} are calculated using equation 2.2 for $i, j = 0, 1, 2, \ldots$.

$$M_{ij} = \sum_{x} \sum_{y} x^i y^j I(x,y) \tag{2.2}$$

The central moments can be defined using equation 2.3. Where $\bar{x} = \frac{M_{10}}{M_{00}}$ and $\bar{y} = \frac{M_{01}}{M_{00}}$ are components of the image centroid, both derived from the raw moments.

$$\mu_{ij} = \sum_x \sum_y (x-\bar{x})^i (y-\bar{y})^j I(x,y) \tag{2.3}$$

The central moments mu_{ij} of any order are invariant with respect to translations. When the central moments is divided by a properly scaled zero-th central moment the result ($\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1+\frac{i+j}{2}\right)}}$) is both invariant to translation and scale, where $i+j \geq 2$. Based on this Hu seven moments can be derived, which are invariants with respect to translation, scale, and rotation.

$$I_2 = (\eta_{20}$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 +$$

$$I_4 = (\eta_{30} + \eta_{12})^2$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 -$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 -$$

Grayscale histogram features are commonly used in image analysis and classification by representing the grey level distribution as a histogram. For example, statistical moments of the grey scale histogram are used as features for the classification of fish species in Spampinato et al. (2010). In this thesis, greyscale histogram features are used as texture to complement Gabor features, which gave information about the spatial arrangement of intensities in the bird species video. They can be used online and have been used by many content-based retrieval systems as features for classification

Gabor wavelets have been applied to many feature extraction problems (Huang et al., 2010; Ou et al., 2010; Parvin et al., 2012; Spampinato et al., 2010) due to its salient visual properties such as spatial localisation, frequency characteristics and orientation.

2.3.2 Motion Features

Curvature scale space (CSS) is rotation and translation invariant and has been shown to be effective in distinguishing object trajectories by their concave and convex shapes (Mokhtarian et al., 1996; Beyan and Fisher, 2013b; Mai et al., 2010; Bashir et al., 2006). They have also been shown to be robust in the presence of noise. Mai et al. (2010) have shown that CSS can be effective for matching and recognizing shapes which are distorted by affine transforms, including translation, rotation, and scaling.

Li et al. (2006) used **turn-based features** to form trajectory directional histogram, which was used to describe the statistic directional distribution of vehicle trajectories. The features were then used for categorisation, achieving an overall accuracy of 99%. Recently, Beyan and Fisher (2013a) have used turn-based features as part of a large feature set to successfully categorise abnormal fish trajectories. Turn-based features was used as part of a feature set due to the unpredictable movement made by fish. Birds like fish also have unpredictable movements, thus it is believed this can be used as part of the feature set for classification of species.

Bashir et al. (2006) have showed that using **Centroid Distance Function (CDF)** for categorisation of features yeilds good recognition rates compared with CSS features. Recently, this has also been used by Beyan and Fisher (2013b) as part of their feature set to discriminate abnormal from normal fish trajectories. CDF is an invariant representation of the shape of data (Beyan and Fisher, 2013b; Bashir et al., 2006).

Vicinity features were used by Liwicki et al. (2006) for handwriting recognition, and recently by Beyan and Fisher (2013b) for classifying abnormal fish trajectories. In both cases, the feature set was made up of vicinity curliness, aspect, slope and linearity.

Liwicki et al. (2006), Chen et al. (2010) and Graves et al. (2009) have used curvature features successfully in handwriting recognition to determine the writing direction and have achieved remarkable results.

2.4 Feature Selection and Reduction

Most of the above methods use a large feature set, but this may affect the performance of the classification model (Hall, 1999; Yu and Liu, 2003) and may contain redundant or irrelevant features, which may lower recognition rate. To improve the performance of such models, feature selection or reduction algorithms have to be applied. Feature reduction techniques are used to transform the feature set in a high-dimensional space to a few dimensional space. The data transformation may be linear, as in principal component analysis (PCA). On the other hand, feature selection is used to choose a subset of the complete set of input features so that the subset can predict the output with accuracy.

The most widely used methods in computer vision for reducing features and improving performance is PCA. PCA has been used successfully by Rodrigues et al. (2010) and in the Fish4Knowledge project (Beyan and Fisher, 2013b; Beyan, 2015) to reduced trajectory features. For bird species, PCA has been applied by Gavves et al. (2013); Zhang et al. (2015) to reduce dimensionality. The benefits of this method are that it reduces training and test times and defying the curse of dimensionality to improve prediction performance. There are also other methods that have been used successfully to reduce the dimensions of *face datasets*, for example, linear discriminant analysis (LDA) used by Martínez and Kak (2001); Yang and Yang (2003). Even though it has been shown by Martínez and Kak, that PCA outperforms LDA, PCA is still widely used in computer vision because of its simplicity.

Another technique for reducing the size of feature sets is feature selection which uses a subset of the features rather than mapping the original features onto a new reduced dimensionality and thus preserving the features (Mladenić, 2006). To select important features for classification a lot of methods have been explored by various researchers (Chandrashekar and Sahin, 2014; Saeys et al., 2007). These can be broadly classified as a filter and wrapper methods. These methods are discussed in the following subsections.

2.4.1 Filter Methods

In filter method, features are ranked in order to determine their importance. A highly ranked feature is considered as the more important feature and lowly ranked, the less important. A number of performance criteria have been proposed for filter method to estimate the goodness of a feature such as Fisher score (Gu et al., 2012), Pearson correlation coefficient (Guyon and Elisseeff, 2003), mutual information (Peng et al., 2005; Yu and Liu, 2003) and ReliefF (Robnik-Šikonja and Kononenko, 2003; Moore and White, 2007). These features have been succinctly described in the following sections.

Fisher score

The Fisher score aims to find a subset of features, such that the distances between data points in different classes are large, while in the same class are small. Precisely, the 'Fishers score Gu et al. (2012) for the i^{th} feature F_i is calculated in Equation 2.4. Where $\bar{x}_{ij}$ and σ_{ij} are the mean and variance of the i^{th} feature in the j^{th} class respectively, n_j is the number of instances in the j^{th} class, and σ_i is the mean of the i^{th} feature. The disadvantage of Fishers score is that it is not good at handling irrelevant and redundant features.

$$F_i = \frac{\sum_{j=1}^{n} n_j(\bar{x}_{ij} - \bar{x}_i)^2}{\sum_{j=1}^{n} n_j \sigma_{ij}^2} \tag{2.4}$$

The top ranked features are then selected after computing the Fisher score but because the scores are independently computed, the features selected are suboptimal. Most importantly, because features are selected not based on their importance with other features, features with low Fishers score but combine effectively with others may be eliminated.

Pearson correlation coefficient

Pearson correlation coefficient is the simplest method for understanding the relationship that exists between the dependent and independent variable. The Pearson correlation

coefficient (PCC) used in Guyon and Elisseeff (2003) ranks features by calculating linear correlations between individual features and class labels in classification.The Pearson correlation coefficient is defined as in Equation 2.5

$$R_i = \frac{cov(X_i, Y)}{\sqrt{var(Xi)var(Y)}} \tag{2.5}$$

where *cov* designates the covariance and *var* the variance. The estimate of R_i is give by Equation 2.6.

$$R_i = \frac{\sum_{k=1}^{m}(x_{k,i} - \bar{x_i})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{m}(x_{k,i} - \bar{x_i})^2 \sum_{k=1}^{m}(y_k - \bar{y})^2}} \tag{2.6}$$

Where x_i stands for the feature value of the i^{th} sample and x is the mean of these feature values. y_i are the labels and y is the mean of all y_i in the sample.

The major advantage of this method is that it is faster and easier to calculate and should be use for ranking features in subset selection Guyon and Elisseeff (2003). It should also be used when there is a high correlation between a feature and the class of the data under consideration Shardlow (2016).

Mutual Information

Mutual Information Peng et al. (2005); Yu and Liu (2003) is one of the popular feature selection methods, as it is computational efficiency and simple to interpret. It is used to calculate the information gain between the i^{th} feature f_i and the class labels C given by Equation 2.7. To determine if a feature is important, there should exist a shared information between the feature and the class.

$$IG(f_i, C) = H(f_i) - H(f_i|C), \tag{2.7}$$

where $H(f_i)$ as in Equation 2.8 is the entropy of f_i and $H(f_i|C)$ as in Equation 2.8 is the entropy of f_i after observing C.

$$H(f_i) = -\sum_j p(x_j) log_2(p(x_j)), \tag{2.8}$$

$$H(f_i|C) = -\sum_k p(c_k) \sum_j p(x_j|c_k) log_2(p(x_j|c_k)) \tag{2.9}$$

The advantage of this method is that it is independent on the classification scheme used but can use any classification scheme to provide error rates. It is also able to treat multi-class cases directly rather breaking them into several two-class problems (Guyon and Elisseeff, 2003). However, the disadvantage is that filters based on mutual information generic feature selection, which is not fine-tuned by the learning algorithm.

ReliefF

Relief select features to separate instance from different classes Kira and Rendell (1992); Robnik-Šikonja and Kononenko (2003); Moore and White (2007). The Relief score of the i^{th} feature S_i is defined by Equation 2.10.

$$S_i = \frac{1}{2} \sum_{k=1}^{l} d(X_{ik} - X_i M_k) - d(X_{ik} - X_i H_k), \tag{2.10}$$

where M_k denotes the values on the i^{th} feature of the nearest instances to x_k with the same class label, while H_k denotes the values on the i^{th} feature of the nearest instances to x_k with different class labels. d is a distance measure. ReliefF was originally designed for two-class problem (Kira and Rendell, 1992) and could not handle noise efficiently. However, a multi-class equivalent has been introduced by Kononenko (1994), which provides an extension to equation 2.10 and also improves it's noise handling capabilities. Due to its inability to eliminate redundancy, recently, Liu et al. (2015) has proposed RS-ReliefF which has been shown to remove redundant data and also improve classification rates on many datasets. Wu and Wang (2015) have also combined sequential forward selection (SFS) with ReliefF (ReliefF-SFS), which was shown to remove the redundant features more effectively than the ReliefF methods alone and also improve the classification accuracy on a music genre dataset. Finally, another method that has shown improvement over

the ReliefF feature selection method is the one proposed by Moore and White (2007), which he called Tuned ReliefF (TuRF). The method symmetrically removes worst performing attributes by re-estimating the ReliefF weights.

The merits of all the filter methods above are that they are classifier-independent and effective with regards to computational cost (Lee et al., 2012). They also scale to large datasets with many features sets, thus performing faster than other methods especially wrapper methods. The major disadvantage of this method is that it may select redundant variables because it does not consider the relationships between variables.

One major work to overcome the disadvantage of the filter method is the correlation-based feature selection method Hall (1999). Hall have determined its performance and accuracy to be similar to wrappers and even under some conditions better. This approach requires no learning algorithms and no threshold settings, as in the case of wrapper, it depends on the learning algorithm and can measure the correlation between variables. The correlation-based approach is based on the assumption that good features are highly correlated with the class but yet uncorrelated with each other (Hall, 1999; Chandrashekar and Sahin, 2014). It eliminates well over half the features, is computationally faster than wrappers and uses pairs of features and subsets (Hall, 1999). The steps involved in the correlation-based approach are:

- Nominal and numeric data are first discretized
- Calculate the feature-class correlation and feature-feature inter-correlations
- Search feature subsets using either best first, forward selection or backward selection algorithms based on their merits which is calculated use Equation 2.11

$$M_s = \frac{k\bar{r_{cf}}}{\sqrt{k+k(k-1)\bar{r_{ff}}}} \tag{2.11}$$

Where M_s is the merit of feature subset containing k features, $\bar{r_{cf}}$ is the mean feature-class correlation (f belong to S) and $\bar{r_{ff}}$ is the average feature-feature inter-correlation.

2.4.2 Wrapper Methods

Wrapper methods use machine learning algorithm such as Random Forest to evaluate the worth of feature subsets by offering a simple and powerful way to address the problem of feature selection (Tang et al., 2014). Wrapper models also achieve better predictive accuracy estimates than filter models. The worth of the selected features in wrapper methods are based on the machine learning algorithm that was applied and they may become unmanageable when the dataset is large, as its performance deteriorates and may become impossible to use. A typical wrapper method was based on variable importance (VI) coming from Classification and Regression Trees (CART), and Random Forests proposed by Breiman (2001). A disadvantage specific to this method is that if the data contains groups of correlated features of similar relevance for the output, then smaller groups are favoured over larger groups (**?**).

Wrapper methods use the machine learning algorithm to evaluate the worth of feature subsets by offering a simple and powerful way to address the problem of feature selection Tang et al. (2014). With a predefined classifier, a wrapper method performs the following steps:

1. Search for a subset of features using any of the search methods succinctly introduced in the following paragraphs.
2. Use the predefined classifier to evaluate the worth of the selected feature subset
3. Repeat steps 1 and 2 until the worth of all features subsets is found
4. select subset with the highest worth

Liu and Yu (2005) have discussed various search methods to find potentially important features for classification. The search usually generates various combinations of feature subsets which are used for evaluation and thus finding the most optimal subset. However, the problem with searching as in the wrapper methods is that, for n features, there are 2^n subsets to be generated, which even with a smaller n an exhaustive search over the feature space is computationally expensive. Feature search in wrapper method thus plays an important role in the selection of important features for classification.

Sequential selection methods also know as greedy search methods start with a feature set and iteratively add or remove features until a termination criterion is met. The are two major types of sequential search methods and these include the sequential forward and backwards search methods (SFS and SBS), which have been introduced in Pudil et al. (1994). The sequential forward selection (SFS) starts with an empty set and adds features until an optimal subset is obtained. The disadvantage of this method is that once a feature is found, it can not be discarded so there is the potential of selecting features that may be sub-optimal. Unlike the greedy forward search, the sequential backwards search (SBS) starts with the entire feature set and then removes features until an optimal feature set is obtained. This method, however, requires more computation than sequential forward selection method. In addition, the sequential forward selection method is widely used, this is because of it's simplicity and speed (Pudil et al., 1994). Typical improvements to the SFS algorithm, known as the Sequential Forward Floating Selection (SFFS) and Sequential Backward Floating Selection (SBFS), proposed by Pudil et al. (1994), were shown to be faster and improve classification rates when compared with the SFS and SBS. SFS and SFFS have been successfully applied to hand-written recognition datasets (Schenk et al., 2009) and have been shown to yield similar results on the hand-written recognition datasets. A more recent method which is an improvement on SFFS is the Adaptive Sequential Forward Floating Selection (ASFFS) algorithm (Nakariyakul and Casasent, 2009), which show slightly better results but requires more computational time when compared with SFFS. Nakariyakul and Casasent show that based on 15 selected features SFFS executed in 6.59 seconds while ASFFS took 182.89 seconds.

Wrapper models obtain better predictive accuracy estimates than filter models but as mentioned already accuracy may also be comparable to the correlation-based feature selection method proposed by Hall. The worth of the selected features in wrapper methods are based on the machine learning algorithm which was applied. With large datasets containing many features, wrapper becomes unmanageable, it performance deteriorates considerably and at times may become impossible to use at all. Since the search in the wrapper method takes a longer time to execute, a lot of methods have been developed

to find optimum sets of features. Some of these have been succinctly introduced in the following sections.

2.5 Classifying Imbalanced Datasets

Machine learning (ML) techniques have been applied in many real-world domains such as credit scoring, biomedical, fish behaviour, bird classification, bat classification and other ecological studies. However, with most of these applications, the required samples for training each class are rare and difficult to collect. Which is fundamentally caused by the intrinsic rarity of the cases or by limitations on data collection process such as high cost or privacy problems (Lee et al., 2009). Datasets with varying sample sizes per class are usually referred to as imbalanced datasets. The class with larger samples is usually referred to as the majority class, and the one with the smaller samples is the minority class.

However, most machine learning algorithms tend to be biased towards the majority class (Ramyachitr and Manikandan, 2014; Lee et al., 2009), especially where there are overlaps in data samples between the classes, which tend to significantly decrease the classification accuracy (Lee et al., 2009). They sometimes fail to classify imbalanced data because the classification error in the majority class dominates the classification error in the minority class (Ramyachitr and Manikandan, 2014). Improving the classification accuracy in the minority class requires imbalanced dataset techniques, which are broadly classified into: re-sampling; Boosting and Bagging; and cost sensitive and algorithm based methods. Resampling techniques are further divided into random oversampling and undersampling. Even though Boosting and Bagging methods are based on random oversampling and undersampling methods, it has been discussed separately in this section. Among the algorithmic method is active learning and those methods that are machine learning dependant (i.e. those where the machine learning algorithm has been modified slightly to handle imbalanced data). A succinct review of these methods is presented in this section.

2.5.1 Resampling Methods

Resampling is the most common technique used in tackling the imbalanced dataset problem. In most cases, either sample from the majority class are eliminated to match the size of the minority class or samples are randomly generated for the minority class to match the size of majority class. There have been various data preprocessing methods (Known as data-level methods) proposed to alleviate this problem. These include re-sampling techniques such as oversampling and undersampling, which have been succinctly reviewed in this section.

Random Undersampling (RUS)

Random undersampling is a technique that aims to balance class distribution by randomly eliminating the majority class samples Ramyachitr and Manikandan (2014) in an attempt to overcome the imbalanced datasets problems faced by most machine learning algorithms. Random undersampling, however, has a major drawback, it can discard potentially useful data that could be important for the training process (Kotsiantis et al., 2006; Yap et al., 2014). Yap et al. (2014) demonstrated that random undersampling is easier to implement and is considered as one of the most effective resampling method Mollineda et al. (2007) to date.

Hitherto, various methods have been developed to overcome the drawback of the random undersampling technique. However, most of these techniques are comparable to RUS in terms of correct classification rates. Kubat et al. (1997) balanced class samples by removing noisy samples from the majority class. This was achieved by first eliminating borderline samples of the majority class using Tomek links, as they were considered as noisy data, and secondly, by eliminating redundant data from the majority class samples. They were able to show that removing redundant samples did not affect the classification accuracy but removing the borderline samples improved accuracy. Another undersampling technique is the nearest neighbour cleansing rule, which removes majority class samples whose class label differs from the class of at least two of its three nearest neighbours (Batista et al., 2004). García et al. (2006) introduced an evolutionary undersampling

technique, which outperformed previously proposed methods such as the nearest neighbour method by Batista et al. (2004). A more recent paper by García and Herrera (2009) contrasted the results using non-parametric statistical procedures to show that evolutionary undersampling outperforms non-evolutionary models when the degree of imbalance increases.

Random Oversampling (ROS)

In random oversampling techniques, samples from the minority class are randomly duplicated to match the number of samples in the majority class or additional samples are synthetically generated and added to the minority class to match the number in the majority Ramyachitr and Manikandan (2014). The merit of oversampling is that no sample from the original training set is lost since all minority and majority samples are kept. However, oversampling can lead to overfitting problems, long training time and increase memory to hold training data. This is because the number of data used in training is much larger than the number of original patterns Lee et al. (2009); Batuwita and Palade (2013).

Various techniques have been proposed to overcome the disadvantage of random oversampling. These include synthetic minority oversampling technique (SMOTE) proposed by Chawla et al. (2002). This is an oversampling technique, in which the minority class is oversampled by creating synthetic samples rather than oversampling with replacement. The synthetic samples are generated in a less application-specific manner, by operating in "feature space" rather than "data space" (Chawla et al., 2002; Kotsiantis et al., 2006). Chawla et al. believes that oversampling becomes more effective when performed using synthetic samples instead of the true instances.

Even though SMOTE has gained popularity, they sometimes generate wrong and unnecessary samples. Because of this problem, other methods have been developed around SMOTE which has achieved significant performance compared with SMOTE. This includes borderline-SMOTE, where only minority samples near the borderline are oversampled. The disadvantage of this technique is that the synthetic data generated are all based on the border line data and thus not representative of the entire data. Han et al. (2005) argued that the samples on the borderline are normally misclassified and therefore

synthetic data should be generated from this. A more recently developed technique based on SMOTE is MWMOTE (Barua et al., 2014), which generates synthetic samples for the minority classes from weighted informative minority class samples and has been shown to be more effective than SMOTE. However, all these oversampling techniques increase the computational cost of the learning algorithm. Another study by Batista et al. (2004) show that it is more convenient to apply oversampling techniques when the dataset has very few minority samples.

2.5.2 Boosting and Bagging Methods

There are other resampling techniques based on boosting and bagging that extract multiple subsets of the same size as the minority class randomly with replacement from the majority class. For example, given a training set S_k of size k, these methods generate n new training sets S_i each of size i. Where S_i is uniformly sampled from S_k with replacement. This meants that some observations may be repeated in each of the n new training sets. The n training sets are then used to form multiple classifiers and majority voting scheme is applied.

Barandela et al. (2003) proposed a method that randomly undersampled the dataset in each bagging iteration, by keeping all minority class instances in each iteration. This technique is based on Bootstrap aggregation (bagging) originally proposed by Breiman (1996). Instead of using one classifier, an ensemble of classifiers was used. The idea is to train each of the individual components of the ensemble with a balanced learning sample. Working in this way, it is possible to appropriately handle the difficulties of the imbalance, while avoiding the drawbacks inherent to the oversampling and undersampling techniques (Barandela et al., 2003). There is no need to drop important samples as in the case of undersampling, as all samples are used, and there will be no need to generate synthetic samples or duplicate random sample, as in the case of oversampling.

Another bagging technique that is based on SMOTE is SMOTEBagging Wang and Yao (2009) that uses SMOTE in each iteration. In SMOTEBagging the new data set contains two times the number of majority class instances, the first half of the data set is a bootstrapped replica while the second is obtained via SMOTE and random oversampling.

In the same paper, Wang and Yao proposed other two methods based on undersampling and oversampling, which he referred to as UnderBagging and OverBagging. Whilst UnderBagging is created by randomly undersampling the majority classes, OverBagging is performed by randomly oversampling the minority classes (Wang and Yao, 2009). Comparing the three methods, the author concluded that SMOTE injects diversity into the ensemble system in most cases, and improves its overall performance compared with undersampling and oversampling.

In order to resolve sample class-imbalanced problems, Yao et al. (2013) recently used an improved Random Forest technique. This technique used SMOTE to deal with the imbalanced dataset problem, splitting the majority class randomly based on the number of samples in the minority. The results were then permuted to form multiple Random Forest classifiers. Their method was tested on five groups UCI Lichman (2013) machine learning datasets and experimental results show that proposed algorithm could improve recall while precision remains the same.

Apart from bagging, boosting techniques have also been used recently to resolve the imbalanced dataset problem. EasyEnsemble proposed by Liu et al. (2009) forms subsets of the majority class, which are used to train a "learner", and then combines outputs of the learners using a voting technique. This performs similar to UnderBagging, but instead of training a classifier for each new bag, each bag is trained using AdaBoost. In the same paper, Liu et al. also proposed a BalanceCascade technique in which learners were trained sequentially and in each step of the sequence, those samples that classify correctly were removed and never considered in other sequences. Liu et al. (2009) have shown this to outperform most existing class-imbalance learning methods whilst having approximately the same training time as the undersampling technique. SMOTEBoost Chawla et al. (2003) introduces synthetic minority class instances using SMOTE algorithm. Weights are assigned to new instances created, which are proportional to the total number of instances in the new dataset. All these algorithms claimed better performance than the original versions of Random undersampling, but still add more computational complexity. RUSBoost Seiffert et al. (2010) removes instances from the majority class in

each iteration of AdaBoost using the random undersampling technique, which provided a faster alternative to SMOTEBoost and yet is comparable in terms of performance.

2.5.3 Learning Algorithm Modification Methods

All the methods mentioned above operate on the data to resolve the class imbalance problem; however, there are methods based on modification of the learning algorithms. These methods were introduced because re-sampling techniques suffer from information loss and overlapping data.

The most common algorithm modification method is the active example selection, developed to resolve the problems of re-sampling mentioned above. Active example selection (AES) technique proposed by Lee et al. (2009) starts with a small balanced subset of training examples and then uses Naive Bayes classifiers to select and add informative examples. Experiments by Lee et al., show that when Naive Bayes is used with this technique, it performs better than sampling methods and cost-sensitive methods. The problem with this method, however, is that slight changes to the initial training data may change the output model (Oh et al., 2011). Due to this problem, Oh et al. developed another method, based on ensemble learning, which they called ensemble active example selection (EAES). This method combines multiple models and uses them as a committee for decision making. Even though ensemble methods are computationally expensive, this method increases the prediction performance of a single model as in the cases of AES.

Recently, various learning algorithms have been modified to resolve the class-imbalanced problem. Tang et al. (2009) tackled the class-imbalance problem by modifying support vector machines (SVM) and incorporating different heuristics (cost-sensitive learning, and oversampling and undersampling) in the SVM modelling. They considered four modifications and of the four, SVMs-repetitive undersampling algorithm (GSVM-RU) was the most effective and efficient (extracting much fewer support vectors, thus speeding up predictions). Again Zhang et al. (2014), modified the SVM classifier to be deal with the class-imbalance problem. Zhang et al. use standard SVM to obtain an approximate hyperplane before using a scaling kernel function, whose parameters were

calculated using chi-square test and weighting factors. Their proposed method was compared to standard SVM, GRNN, PNN and ELM and the results shows that the proposed method outperformed the others in three of the four datasets used.

Apart from SVM, other classifiers have also been modified to deal with the class-imbalanced problem including the Naive Bayes classifier. Recently, Sobran et al. (2013) demonstrated the Naive Bayes classifier produces better classification accuracy on datasets that are nearly balanced, but struggles when the datasets involved are class-imbalanced. They thus proposed that the Naive Bayes classifier requires improvements and modifications before it can be used in solving imbalanced dataset problems (Sobran et al., 2013). Frank and Bouckaert (2006) proposed a Naive Bayes method that adjusts attributes priors and showed that it can significantly improve the area under the ROC curve. This was adapted by Chawla et al. (2002), making it cost-sensitive by varying the priors of the minority class. Thus, when compared with other classifiers, it only performed better with the Pima dataset, which is the least skewed.

2.6 Multi-Class classification

In machine learning, multi-class classification is the problem of classifying instances into one of the $N > 2$ classes. Thus, in multi-class classification, each of the training data belongs to one of the $N > 2$ different classes and the aim is to develop a function, which will correctly predict the class to which a test sample belongs.

Whereas some classifiers like the Naive Bayes, Decision-tree, Random Trees, Random Forest, naturally permit the use of more than two classes, others like the SVM and the KNN are binary classifiers. A binary classifier is one which can classify instances into one of two classes.

Decision-tree, random trees and random forest algorithms can be easily generalised to handle multi-class classification problems. This can be done by assigning each leaf of the tree to one of the N classes and selecting internal nodes to discriminate among these classes (Dietterich and Bakiri, 1995). According to Li et al. (2004), Naive Bayes also naturally permits multi-class classification. The classifier is based on the Bayes rule

assuming conditional independence between classes. Given an observation, the classifier estimates the conditional probabilities of classes using the joint probabilities of samples and classes and the observation is classified as one of the N classes based on the conditional probabilities.

Binary classifiers can be turned into a multi-classifier using some strategies. Support vector machine (SVM), proposed by Vapnik and Vapnik (1998), had been an excellent binary classifier. It searches for the hyperplane that separates the two classes of data with the largest margin. However, they have been used in Li et al. (2004) and Tax and Duin (2002) for multi-class classification using the one-versus-one and one-versus-all strategies.

In the one-versus-one strategy, a classifier is trained for each pair of classes, thus resulting in $k(k-1)/2$ independently binary classifiers. In this approach, to predict the class for a test instance, each of the $k(k-1)/2$ classifiers are used, and a voting technique, typically, majority vote (class with the maximum number of votes) is applied to determine the class of that test instance. In the one-versus-all strategy, a single classifier is trained per class, unlike the one-versus-one where k classifiers are trained per class. The training samples from the k^{th} class are used as positive samples and training samples from the other $k-1$ samples are used as negatives. Again, unlike the one-versus-one strategy, where the classifiers outputs are class labels, this requires the output of classifiers to be real-valued confidence score. The class with the highest score is assigned to the test instance. The difference between these two methods is their computational cost, as the one-versus-one needs $k(k-1)/2$ classifiers and the one-versus-all requires only k classifiers.

2.7 Machine Learning Algorithms

Classification is the process of identifying the class a test instance belongs to, given some training set whose class membership are known a priori. In machine learning and statistics, classification can either be binary (having two classes) or multiclass (having more than two classes). Most of the earlier classifiers were developed to be used for binary

classification. Until recently, most of these classifiers only support multiclass classification by implementing multiple binary classifiers.

There is an exhaustive list of classification methods used for animal species. The performance of which depends on the characteristics of the dataset used. There is no ideal classifier that works well on all problem domains but to review all classifiers is impossible given the scope of this thesis. Therefore four standard classifiers, including the Naive Bayes, Support Vector Machines (SVM), Random Tree and Random Forest, which have been briefly reviewed in the following sections.

2.7.1 Support Vector Machines Classifier

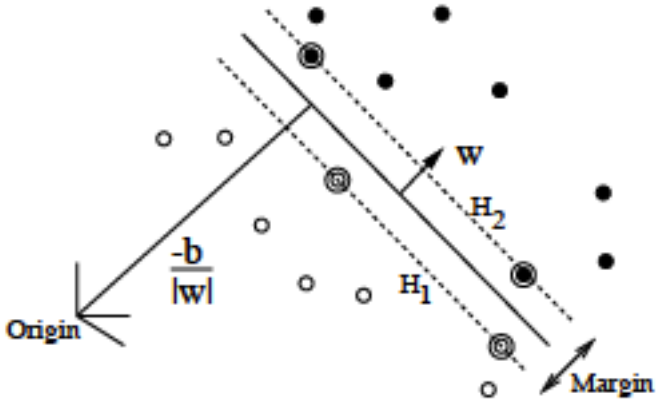
A support vector machine (SVM) is a supervised machine learning algorithm which is used for classification but sometimes for regression. SVM constructs a hyperplane or a set of hyperplanes in a high-dimensional space for classification or regression. The best hyperplane is used to separate instances of the test set. This hyperplane is the one that has the largest distance to the nearest training data point of any class, called the margin. The larger this margin the lower the generalisation error of the classifier. The hyperplane function can take a vector input and output it's class membership (Burges, 1998; Frezza, 2013).

Given a training dataset of n points and two classes, of the form $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$. Where y_i represents a positve ($y_i = 1$) or negative ($y_i = -1$) class. Each point in $\vec{x}_i$ belongs to either the positve or negative class of y_i. The aim of the SVM classifier is to find the "maximum-margin hyperplane" that divides the group of points into positives and negatives. The "maximum-margin hyperplane" is defined such that the distance between this hyperplane and the nearest point $\vec{x}_i$ from either group is maximized. The hyperplane can be written as set of points satisfying Equation 2.12

$$\vec{w} \cdot \vec{x} + b = 0, \tag{2.12}$$

Where $\vec{w}$ is the normal vector to the hyperplane and $\frac{b}{\|\vec{w}\|}$ determines the offset of

Figure 2.3: Linear separating hyperplanes for the separable case. The support vectors are circled and H_1 and H_2 are two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible. Adapted from Burges (1998)



the hyperplane from the origin along the normal vector $\vec{w}$. If the training data are linearly separable, we can select two parallel hyperplanes ($H_1 and H_2$) that separate the two classes of data, so that the distance between them is as large as possible. The region bounded by these two hyperplanes (Equations 2.13 and 2.14) is called the "margin", and the maximum-margin hyperplane is the hyperplane that lies halfway between them (see Figure 2.3). This constraint states that each data point must lie on the correct side of the margin (see Equation 2.15).

$$\vec{w} \cdot \vec{x} + b = 1 \tag{2.13}$$

$$\vec{w} \cdot \vec{x} + b = -1. \tag{2.14}$$

$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \quad \text{for all } 1 \leq i \leq n. \qquad (1) \tag{2.15}$$

In non-linear problems, SVM uses a function called the kernel to transform the input non-separable data to a separable input space. Kernels used in this case include Polynomial, Radial Basis Function (RBF) and Hyperbolic tangent. For the RBF, the kernel used

is $k(\vec{x_i},\vec{x_j}) = \exp(-\gamma\|\vec{x_i}-\vec{x_j}\|^2)$, for $\gamma > 0$. Sometimes parametrized using $\gamma = 1/2\sigma^2$. Whereas the Hyperbolic tangent uses the kernel, $k(\vec{x_i},\vec{x_j}) = \tanh(\kappa\vec{x_i}\cdot\vec{x_j}+c)$, for some (not every) $\kappa > 0$ and $c < 0$.

However, the effectiveness of SVM depends on the selection of these kernels, their parameters, and soft margin parameter C. Usually, the best choice of C and γ yield the best correct classification rates. The best values for these parameters can be obtained using grid search (Hsu et al., 2003) usually, with an exponential growing sequence, for example $C \in \{2^{-5},2^{-3},\ldots,2^{13},2^{15}\}$; and $\gamma \in \{2^{-15},2^{-13},\ldots,2^{1},2^{3}\}$. However, a complete grid search is time-consuming. The approach used by Hsu et al. (2003) can be applied, by first performing a coarse grid search. When a good region on the grid is found, a fine grid search is performed within that region, to reduce the search time for both parameters.

SVM have been used extensively for recognition of human actions and emotions (Schüldt et al., 2004; Meng et al., 2007), and in large-scale image classification (Lin et al., 2011; Akata et al., 2014). Recently, an SVM was used for fine grain classification of objects by parts (Berg and Belhumeur, 2013; Zhang et al., 2015) and by alignment (Gavves et al., 2013). Ge et al. (2015) used a one-vs-all SVM for classifying species, and Hall and Perona (2015) used it for fine grain classification of pedestrians. A more recent work by Berg et al. (2014) used SVM for large scale fine grain categorisation of 500 US bird species. Díaz-Uriarte and De Andres (2006), compared the performance of SVM to Random Trees (RT) and Random Forest (RF) and found out that the classification accuracy decreases as the training set increases. However, when used as a binary classifier, the classification accuracy was better than the RT and RF classifiers. SVM moves the problem of over-fitting from optimising the parameters to model selection. The kernel models can also be quite sensitive to over-fitting the model selection criterion (Cawley and Talbot, 2010). Finally, Techniques such as SVM have excellent performance for balanced data but may fail when applied to imbalanced datasets (Anand et al., 2010).

2.7.2 Naive Bayes Classifier

Naive Bayes (NB) is a method use in constructing classifiers models that assigned a class label drawn from a finite set to an instance, which is represented as a feature vector of

values. This classifier assumes that the values of any feature set are independent of each other given a particular class label. For example, a bird may be considered as a Chaffinch, if its size is of a robin or smaller, found in farmlands and feathers are red, black, grey and yellow. A Naive Bayes classifier considers each of these features as contributing independently to the probability that this bird is a Chaffinch, regardless of any possible correlations between the feather colours, size, and habitat features. One can work with the Naive Bayes method without using Bayesian probability or any Bayesian methods, as parameter estimation can use the maximum likelihood method (Hald, 1999).

The Naive Bayes is a conditional probability model. When given an instance to classify, which has features represented by a vector $x = (x_1, \ldots, x_n)$ of n independent features, the NB assigns the instance some probabilities $p(C_k|x_1, \ldots, x_n)$ for each of K classes C_k. The above formulation is flawed, especially when the number of feature n is large and the probability table becomes infeasible. This model can, therefore, be reformulated using the Bayes' theorem. The conditional probability can be decomposed as in Equation 2.16.

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \tag{2.16}$$

Where $p(C_k|x)$ is the posterior probability of class c given attribute x, $p(C_k)$ is the prior probability of class C, $p(x|C_k)$, is the likelihood, which is the probability of feature x given class C and $p(x)$ is the prior probability of the observed feature.

Now, constructing a classifier from the probability model, we assume conditional independence (that is each feature is conditionally independent of every other feature) and combine this model with a decision rule. The Bayes classifier is thus a function that assigns a class label $\hat{c} = C_k$ for some k as in Equation 2.16.

$$\hat{c} = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i|C_k) \tag{2.17}$$

However, dealing with continuous data, one common practice is to assume that continuous values associated with each class are distributed according to a Gaussian distribution. The data is segmented according to class c and the mean μ_c and standard deviations

σ_c^2 are computed. The probability distribution of x_i given a class c, $p(x = x_i|c)$ can be computed using the normal distribution equation (see Eqn. 2.18).

$$p(x = x_i|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(x_i-\mu_c)^2}{2\sigma_c^2}} \tag{2.18}$$

Naive Bayes is often used as a baseline learning algorithm (Rennie et al., 2003) and has proven effective in many practical applications, including text classification, medical diagnosis, and systems performance management (Domingos and Pazzani, 1997; Hellerstein et al., 2000) and for Recognising human facial expression and emotion (Sebe et al., 2002) due to it being fast, even on large datasets.

Naive Bayes classifier only requires a small number of training data to estimate the parameters necessary for classification (Zhang, 2004). It is well known that Naive Bayes performs well in classification, but its probability estimation is sometimes poor (Zhang and Su, 2008). Zhang and Su have also shown that the Naive Bayes classifier normally favours classes which have a larger number of samples, and also does not work well on noisy datasets.

2.7.3 Random Decision Tree Classifier

A Random Decision Tree (RDT) classifier proposed by Fan et al. (2003) is a non-parametric supervised machine learning algorithm which is used for classification. Unlike classical decision tree methods, RDT constructs of multiple decision trees randomly at each node and does not need any purity function checks such as Gini Impurity (Mehta et al., 1996) or Information Gain (Fan et al., 2003). As a requirement, the number of trees needs to be set prior to training and classification. Random decision trees may introduce large variation in accuracy, Fan et al. (2003) recommended running the test at least five times and for up to a total of 100 trees; however, with most datasets, ten trees are enough to achieve good accuracy. The algorithm first starts by building a structure of N random trees without data and then the training data is scanned one at a time, in order to update statistics at each node of the random trees. The structure of each tree can then be simplified after training without altering the logic of the trees. During prediction, the posterior

probability is estimated by averaging the probabilities from multiple random trees.

Various techniques have been developed to improve the performance of random decision trees classifiers. These include a differentially private decision tree ensemble algorithm, developed by Jagannathan et al. (2009). This has been shown experimentally to yield good prediction accuracy on three UCI Machine Learning Repository (Lichman, 2013), namely Nursery, Congressional Voting Records and Mushroom datasets. Zhang et al. (2010), have developed a technique that uses random decision tree for multi-label classification. This was experimentally shown to be more computationally efficient than state-of-the-art methods, using large datasets. It was also observed that the RDT classifiers had 10% more correct classification rate when compared to the state-of-the-art, including Naive Bayes and K-Nearest Neighbor(KNN) on more than five datasets. Many attempts have been made (Oberg et al., 2012; Jagannathan et al., 2009; Zhang et al., 2010) to improve the computational efficiency of decision trees. Recent work by Oberg et al. (2012) attempted to increase the speed of random decision tree for body parts recognition, using Microsoft Kinect and Field-Programmable Gate Array (FPGA). Random decision tree classification is powerful but is a computationally expensive machine learning algorithm. Supporting random decision trees raises significant computing challenges, particularly for high duty-cycle embedded applications (Oberg et al., 2012). Semi-random multiple decision trees (Hu et al., 2007) is a method based on the random decision tree, which has been used to the classifying data stream and shown to outperform state of the art, in terms of computational complexity and accuracy.

The merits of using random decision trees usually lie in its ability to use minimal memory and also it using less time during the training phase (Zhang et al., 2010). RDT has been demonstrated by Fan et al. to achieve higher accuracy on most datasets when compared with more expensive combination techniques such as boosting, bagging, meta-learning and MetaCost.

2.7.4 Random Forest Classifier

There has been a lot of interest in "ensemble learning" methods (Breiman, 1996; Liaw and Wiener, 2002; Freund et al., 1996; Schapire, 1990) that generate many classifiers and use

them to predict a class by aggregating their results. The two most popular are boosting (Schapire, 1990) and bagging (Breiman, 1996) of classification trees.

Bagging predictors was proposed by Breiman (1996) to improve the classification by combining classifications of randomly generated training sets. The training sets are generated using bootstrapping and aggregation is used in combining classifications of the training set. Usually, in bagging, the aggregation is done by either averaging the results or using majority votes (mode). Bagging Trees was shown by Breiman (1996) to outperform other traditional classifiers like a single Tree and Nearest Neighbor classifiers. In particular, Bagging Tree outperformed Nearest Neighbor classifier on all six datasets used whereas a single tree only outperformed Nearest Neighbor on three out of the six classifiers. One major problem with "Bagging" is determining the number of bootstraps to use, which in the case of Breiman was 50. However, Breiman showed that misclassifications reduced as the number of bootstraps increase. Another limitation of this method is that when the number of bootstraps is not large enough, some samples are left out of the training set. Another ensemble method that depends on aggregation is Boosting. This was proposed by Schapire (1990) and is based on the concept that the training set class is weakly learnable if the learner can produce a hypothesis that performs slightly better than random guessing. In Boosting, a series of "weak" classifiers are built, each being trained on a dataset in which classes misclassified by the previous classifier are given more weight. and all the classifiers are then weighted according to their success and their outputs are combined using majority vote. The most common boosting classifier is AdaBoost (Freund et al., 1996), which stands for "adaptive boosting". It combines the outputs of many "weak" classifiers to increase classification accuracy. The advantage is that weak classifier can be very simple to implement and computationally inexpensive (Friedman et al., 2001).

Random forest proposed by Breiman (2001) is emerging as a very popular classifier. This is an extension of the bagging method and was developed to improved classification performance when compared with using bagging and boosting classifiers such as AdaBoost. It is based on trees but each node is split using the best among a subset of

predictors randomly. It aims to reduce the variance of the individual tree by randomly selecting many trees from the dataset and averaging their prediction output. This method is based on bagging and random decision trees, discussed in section 2.7.3. Random Forest classifier Breiman (2001) introduces a random permutation into the learning process, in order to produce multiple decision trees from a single dataset (thus forming a "forest"). Aggregation techniques, such as majority voting, are then used to combine the predictions from all of the trees. This method combines Breiman's "Bagging" (Breiman, 1996) whilst injecting random perturbations into the feature selection, for building a collection of decision trees with a controlled variance. The dataset is split into a training set and a test set, known as the *out of bag* cases, in a Random Forest classification problem.

Random Forest classifiers have been applied in many applications, including bioinformatics (Lee et al., 2005; Díaz-Uriarte and De Andres, 2006; Statnikov et al., 2008), and ecology for the classification of vascular and non-vascular plants and for vertebrates Cutler et al. (2007). Lee et al. (2005) extensively analysed the performance of many classifiers (21 classifier methods) in analysing microarray gene expression data (7 datasets) and found out that Random Forest was the most successful classifier on these datasets. However, another comparative study, though not extensive, was carried out by Statnikov et al. (2008) to compare SVM and Random Forest for disease samples classification. Their results show that the SVM classifier outperforms Random Forest in this case. Apart from applying Random Forest directly to the classification of genes, the study by Díaz-Uriarte and De Andres (2006) had looked into gene feature selection prior to classification. They used Random Forest to select genes for classification and found out that Random Forest has comparable performance to other classification methods, including Linear discriminant analysis (LDA), KNN, and SVM when feature selection was carried out. In ecology, Cutler et al. (2007), compared the accuracy of Random Forest with other commonly used statistical classifiers such as LDA, Classification trees and Logistic regression for classifying plant species and bird habitat datasets. They observed that Random Forest had higher classification accuracy.

The advantage of this method is that it improves the randomness in the bagging

method, and in addition, it has only two parameters (the number of variables in each subset and the number of trees). It is also not sensitive to these values (Liaw and Wiener, 2002). It has been shown in (Liaw and Wiener, 2002; Díaz-Uriarte and De Andres, 2006) that the Random Forest classifier compares quite favourably with SVM and on some datasets, it may even outperform it. Cutler et al. (2007) also mentioned that Random Forest classifier has a generally high classification accuracy, and is able to model complex interaction among predictor variables. Finally, it retains the many benefits of Random Decision Trees, as well as achieving better classification accuracy by using random subsets of variables, majority voting, and bagging samples (Qi, 2012).

2.7.5 Choice of Classifiers

Different researchers have reported differing results, and the choice of one classifier over the other depends to some extent on the dataset used. For example, It was shown in Huang et al. (2003) and Karim and Rahman (2013) that in their case Naive Bayes Classifiers has similar accuracy to various variants of tree classifiers, whereas Huang and Ling (2005) showed that Naive Bayes classifier is significantly better than decision trees. Similarly, research carried out by Lee et al. (2005) in analysing microarray gene expression data found Random Forest to have better accuracy than other classifiers, including SVM, whereas Statnikov et al. (2008) found SVM to have better accuracy than Random Forest.

Since it is difficult to determine which classifier works best for which dataset, this research have experimented with four different classifiers (Naive Bayes, Random Decision Tree, Random Forest and Support Vector Machines) and evaluated their results. SVM classifier was have selected because of its recent support for multi-class classification, and it has been used for most computer vision applications for fine-grained classification (Krause et al., 2015; Gavves et al., 2013). Naive Bayes classifier was chosen because it has been used extensively and had no parameter values to tune, thus making its implementation relatively easy. Random Decision Tree was used because they have also been shown to have generally higher accuracy (Fan et al., 2003) and are less computationally expensive (Zhang et al., 2010). Finally, Random Forest was chosen because it has been shown to compare favourably with other classifiers and sometimes performs better (Liaw

and Wiener, 2002; Díaz-Uriarte and De Andres, 2006) by achieving good classification accuracy using random subsets of features, bagging samples and using majority voting (Qi, 2012).

2.7.6 Statistical Significance of Two Classifiers Over K-folds

There are lots of statistical significance tests used in comparing the performance of two classifiers. The most commonly used are the sign test, Wilcoxon signed-rank test, k-folds paired t-test, r-times k-folds paired t-test, 5x2cv paired t-test and the McNemar's test. However, little work has considered the statistical significance of two classifiers over a k-fold, where $k > 2$.

One way to compare the statistical significance of the results of two classifiers over k-folds is to apply the k-fold cross-validated paired t-test. The test sets used in this method are independent of each other but the training sets overlap. Results are slightly affected compared to when the training set are completely independent (Dietterich, 1998; Peters et al., 2003). In Peters et al. (2003), a 10-fold cross validation was used and each of the ten results by the two learning algorithms was subjected to a paired t-test. An alternative to the k-fold cross-validated paired t-test is the 5x2cv paired t-test, which Peters et al. showed to eliminate the Type I error exhibited by the aforementioned method and which have been used in (Dietterich, 1998) to compare two classifiers. This method is based on five runs of a two-fold cross-validation. Bouckaert and Frank (2004); Nadeau and Bengio (2003) proposed r-times k-fold cross-validation pair t-test, which is a repeated k-fold cross validation. This method has been used together with the corrected t-test in Corani and Benavoli (2015), to compare two classifiers over a single dataset.

Another method that has been successfully applied to test the statistical significance of two classifiers over a single dataset is the sign test. In Nadeau and Bengio (2003), sign test was applied on a k-fold cross-validation and r-times k-fold cross-validation. However, Nadeau and Bengio recommend running the k-fold cross-validation r-times and averaging over the folds separately for the two classifiers before applying test on the difference of the classifiers. It was recommended that r should be greater than 10, as this results in

reasonable replicability, but the value should also be chosen with computational consideration.

All the above-mentioned methods that are based on paired t-test (apart from 5x2cv paired t-test) can be applied to a 5-fold cross validation. However, they are based on assumptions: the randomness of the samples, the two samples should come from populations with equal variance and the data is required to fit a normal distribution. The sign test and Wilcoxon's signed-rank test can also be applied to a five-fold cross-validation results. The advantage of these two methods over the methods that use paired t-test is that they are non-parametric (the data is not required to fit a normal distribution). In addition, Wilcoxon's signed-rank test is not substantially affected by the presence of outliers Demšar (2006). This test is considered more powerful than the signed test and is a non-parametric alternative to the paired t-test (Fay, 2007). In Fay (2007), to achieve better and comparable results to the paired t-test using a k-fold cross validation, the k-folds were repeated r-times. The results of the *r.k* experiments performed with both learning algorithms were used as input for the Wilcoxon statistical significance tests.

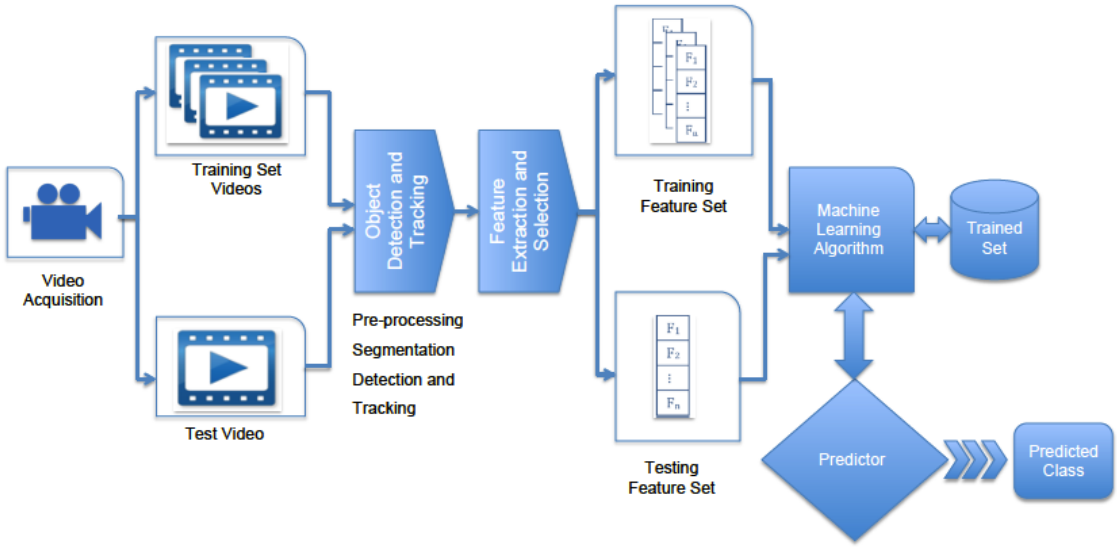
In this thesis, to compare the statistical significance of the results of two classifiers, the Wilcoxon's signed-rank test is applied on r-times k-fold cross-validation results as in Fay (2007). This has been chosen for various reasons: the test is non-parametric; it requires no assumptions; it has been shown in Fay (2007) to produce better and sometimes comparable results to the paired t-test methods; and has been shown to be more powerful than the sign test. Furthermore, the r-times k-fold cross-validation was used with this method as this has been used in Fay (2007) and its been shown in Nadeau and Bengio (2003) to produce better than using the standard k-fold cross-validation.

2.8 Brief Overview of a Video Classification System

Existing video classification systems for species recognition are generally composed of two stages. The first stage (the extraction and training stage) is when videos are labelled, features extracted and feature database formed with labelled species, training of the dataset, which is used to form classification decisions. The second stage (classification

stage), uses new instances of videos and based on the classifier decision procedure tries to identify species. A brief overview of a video classification system have been presented in this section to help the reader understand other chapters of this thesis.

Figure 2.4: Overview of a video framework for classifying species. This figure shows the video acquisition stage, dividing the videos into training and test sets and preprocessing of the videos. It also includes segmentation, detection and tracking of birds from frame to frame. The final stage of this system is the feature extraction, selection and classification.



The first step in the species classification system is a method of obtaining the input video, which consists of a sequence of image frames obtained by using a video camera. The type of video camera that should be used is determined by the type of application being developed. For example, a system that will be used at night will use a thermal imaging video camera and a computer vision application for medical imaging diagnosis, may require high resolution images, which offers a high pixel density and thereby more details about the original scene (Mastriani, 2014). For computer vision systems that require rapid motion, higher frame rates promises better tracking (Handa et al., 2012), but this is at the expense of high computational and storage requirement. In the work presented in this thesis, the decision was made to work with an off-the-shelf high-speed digital camera for two reasons: they capture fine details of flight patterns especially for those species with faster wingbeats and secondly, because the cost of off-the-shelf cameras is relatively cheap, thus offering a cost-effective solution.

Once the video has been acquired, various processing methods can be applied to the image in order to perform different tasks that are required in many vision applications (Moeslund, 2012). Nevertheless, if the video acquired is very noisy, then the intended task may not be realisable, even if some enhancement techniques have been applied. Various factors could lead to extremely noisy videos, which include where the hardware is not properly configured introducing visual artifacts. Normally the goal of pre-processing is to enhance the visual appearance of the video and effectiveness of processing techniques. Extra caution must be put in place to ensure enhancement techniques do not introduce further artefacts or lead to information loss. There are various ways to enhance the quality of a video. One common way is by removing noise in the image. This can be achieved by using mean or median filtering. Another way is to enhance particular features in the image by using Morphological operations such as subtraction, dilation and erosion. Subtraction helps eliminate confusing background details which remain unchanged between image frames. Dilation connect features in videos whilst erosion disconnect features in videos by removing smaller objects.

Once prepocessing has been applied, the detection process will begin with segmentation of objects from the video frames. The simplest way to do this is to make the initial frame of the video the background and then subtract it from subsequent image frames to achieve the foreground image (known as frame differencing). This algorithm is not robust to scenes with many moving objects and those with background models that changes over time. Another method that can be used to segment objects in videos is the Running Gaussian average. Wren et al. (1997) propose fitting a Gaussian probabilistic density function (pdf) characterised by mean μ_t and variance σ_t^2 on the last n frames. the running average computed in order to avoid fitting the pdf from scratch at each new frame time t. This technique has been shown by Wren et al. to have good indoor performance. The disadvantage of this method is that it takes a relatively long time to obtain a good estimate of each pixel's colour covariance. Stauffer and Grimson (1999) provided a method which outperformed the method presented by (Wren et al., 1997) and was able to also achieve good outdoor performance. Their method modelled the values of a particular pixel as a

mixture of Gaussians (MoG) and based on the variance of each Guassian mixture model, they determined which Gaussian may correspond to a background (Stauffer and Grimson, 1999). The disadvantages of this method are that when the number of Gaussian mixture models increase beyond 5, it uses up a lot of memory space (the authors recommended using 3 or 5 Gaussians) and that it is not able to distiguish betwen moving shadows and objects. The method also suffers from slow learning at the beginning when in busy environments. KaewTraKulPong and Bowden (2002) improved the algorithm by Stauffer and Grimson (1999) by utilising different update equations at different phases to enable the system to learn faster even at the beginning. They also introduced a shadow detection scheme into the system. The improved method by KaewTraKulPong and Bowden (2002) produced far better segmentation when used with shadow detection. This adaptive method was then improved by the work in Zivkovic (2004). This algorithm automatically selects the required number of mixture models and fully adapts to observed scenes. When this method was evaluated on the datasets it was found that it is more suitable for real-time applications as the processing time is reduced and the segmentation improved. A complete survey of MoG methods can be found in Bouwmans et al. (2008).

Once the foreground image is segmented from the background, the objects of interest need to be detetected and tracked while in view. Object detection deals with finding instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Object tracking involves determining the spatial-temporal trajectory of the object of interest in the video sequence. The two are closely related as object detection always preceeds tracking and is always required throughout the tracking process. "The aim of an object tracker is to generate the trajectory of an object in the image plane as it moves around a scene" (Yilmaz et al., 2006). Objects can be repesented as points (centriod) or set of interesting points (Shi and Tomasi, 1994; Schmid et al., 2000), geometric shapes or as silhouettes and contours (Yilmaz et al., 2004). Tracking can be confounded by noise in the video, partial or full object occlussion, complex objects shape and scene illumination changes. Object tracking is important for this research work, as it uses motion features. There are various methods used in object tracking, which include

point or centroid tracking, Silhouette Tracking, Shape Matching, and Contour Tracking. When tracking points, various criteria could be used to match an object in one frame to another object in the next frame. One way is to assume the proximity of the object will be within a specified threshold in the next frame. In this approach, feature point correspondence becomes a problem as a feature point in one frame may have many matches in the next (Deori and Thounaojam, 2014). Other methods can predict the position of an object from frame to frame, such as Kalman and particle filters (Arulampalam et al., 2002). Silhouette tracking (Rosenhahn et al., 2005) is employed when tracking of the complete region of an object is required. The goal of a silhouette-based object tracker is to find the object region in each frame by means of an object model generated using the previous frames. This method is known to be computationally efficient but this may deteriorate with occlusion (Deori and Thounaojam, 2014). Contour-based object tracking (Yokoyama and Poggio, 2005; Yilmaz et al., 2004) track contours based on the outlines of the object from one frame to the other. The disadvantage of this method according to Deori and Thounaojam (2014) is that it is highly sensitive to the initialization of tracking, thus making intitial automatic tracking difficult.

After objects have been detected and tracked in the object classification systems, the appearance features are extracted from the object's silhouette in the current frame whilst the motion features are extracted from the object's trajectories over a window of frames. Feature extraction starts from an initial set of measured data and builds derived features intended to be informative, facilitating subsequent learning. When the number of features is large there may be redundant and irrelevant features. Features selection can transform the large feature set into a reduced feature set by eliminating redundancy and improving performance. A summary of feature selection methods can be found in Section 2.4. Finally, after feature selection, classification can be performed. Details of machine learning algorithms for classification can be found in Section 2.7. In machine learning and statistics, classification is identifying which class an observed sample belongs, based on a training set.

2.9 Summary

It has been shown in this chapter that computer vision techniques have been applied not only to birds monitoring and classification, but also to other species. Applications to bats and fish have been reviewed: techniques used for these species are similar to those used for birds, and most researchers studying these species have considered motion features, a technique which is of importance to this thesis. Bio-acoustics techniques have also been used for the classification of bird species, but have clear limitations, as described.

Computer vision techniques used for the classification of bird species from a single frame have been reviewed, and its been noted that these can be divided into those which are part-based and those which are not. Part-based methods have generally gained better classification accuracy, but require some manual input and good-quality images in which body parts are present and identifiable. These manual inputs are usually annotations which identify the bird's parts prior to feature extraction. Decomposing soft-bodied objects with arbitrary poses remains a challenging problem in computer vision. The major drawback with most of the non-part-based method is that just like part-based methods, they make use of the Grab-Cut segmentation method which so still require some manual input.

Motion features have recently been studied for the broad categorisation of a small number of species; for example, to distinguish between birds and bats. Researchers that use this approach rely on the use of autocorrelation, FFT and object similarity-based periodicity to analyse motion patterns. These techniques have been used mainly to remove noise in the motion signal which might be a result of bad segmentation. Motion features have mainly been used with bats and fishes, but all were based on broad categorisation; for example, detecting abnormal trajectories. None of these, however, have considered the use of combined motion and appearance feature for classification of bird species.

Combining motion and appearance features, may lead to a large feature set, which may contain irrelevant and/or redundant features. Feature selection techniques may therefore, be useful to select important features, without having to transform the dataset. These techniques were reviewed broadly as filter and wrapper methods. It was noted that filter

methods are generally faster than wrapper method; while the wrapper method may provide better correct classification performance. The correlation based method proposed by Hall (2000) however, produces classification accuracy which is comparable to most wrapper methods although still being faster.

The next chapter presents a preliminary investigation that uses two proposed modification to the baseline algorithm to estimate wing beat frequencies. For intial investigation this was applied to a video dataset of flying bats.

Chapter 3

Preliminary Analysis of Wing Beat Frequency

The objective of the work presented in this chapter is to develop algorithms to automatically analyse the features of species' flight behaviour, and quantifying wing beat frequency using spectral analysis. Physical characteristics, such as body mass and species, are known to vary with wing beat frequencies (Bullen and McKenzie, 2002; Norberg and Norberg, 2012), and may potentially be used for automated classification. The work presented here uses spectral analysis techniques to quantify wing beat frequencies, using a single imaging device in low-light. Two modified techniques based on bounding box metrics were proposed, and similarity matrices, for measuring periodic and cyclical motion as a 1D time domain signal. These are transformed to the frequency domain using Fast Fourier Transform (FFT) and the techniques evaluated against the baseline algorithm proposed by Cutler and Davis Cutler and Davis (2000), using expert-annotated ground-truth data. Bats were initially used for this study partly because the data was readily available but also because bat motion is of a higher frequency and so is more challenging for analysis.

The rest of the Chapter is organised into the following. In section 3.1, a description of the dataset used for the experiments and a discussion of low-level image processing techniques for dealing with low-light video data of bats are presented, In section 3.2 the experimental setup is detailed and results presented in section 3.3 and finally section 3.4

summarises key conclusions.

3.1 Dataset and Methods

The data used in this research was provided by the Lincolnshire Bat Group. The bats used were rescued and being rehabilitated prior to release. The video sequences were recorded using a Casio Exilim ZR100 recording at 240 frames per second. The IR filter was removed from the sensor of this device to facilitate IR illumination for low-light filming. This section discusses the datasets and methods used in the experiments presented in this Chapter.

3.1.1 Dataset

Three samples were used in experiments performed in this chapter, taken from a single Common Pipistrelle. The videos were taken on different days, with different recorded flight weights. The first sample used was broken into two parts. The first part had 113 frames (about 0.5 seconds of data) and the second part 96 frames (about 0.4 seconds of data). The second sample had 130 frames (about 0.6 seconds of data) and the last sample 508 frames representing about 2.2 seconds of data.

3.1.2 Methods

The proposed method were derived from Cutler and Davis (2000), but with a number of proposed modifications. An investigation shows that a similarity metric based on silhouette shape is not as effective with a target which can change orientation arbitrarily. Therefore, a coarser metric derived from the oriented bounding box fitted to bat's 2D object silhouette was proposed to better capture the periodicity of motion (as well as being computationally less expensive). Further investigation shows that the selection of the dominant frequency proposed by Cutler and Davis is often inconclusive; therefore, two techniques were proposed to replace it. The first technique is called the diagonal selection (DS), which is based on the correlation of the signal with individual components, reconstructed from the peaks of the signals' frequency spectra and the second is the self-similarity technique.

Algorithm 1: Fit an oriented bounding box on a set of 2D points (bat silhouettes)

```
minArea ← 0;
foreach Edge ∈ ConvexHull do
    Orientation ← compute the edge orientation;
    Rotate the ConvexHull using Orientation;
    area ← Compute bounding box area of the rotated convex hull;
    if area < minArea then
        RECTANGLE ← Rectangle of minArea;
    else
    end
end
Return RECTANGLE
```

For each video, the bats' silhouettes (Figure 3.1) using the background Gaussian mixture model proposed by Zivkovic and van der Heijden (2006) was extracted . This method was used as it's been shown (Zivkovic and van der Heijden, 2006; Bouwmans et al., 2008) to be more suitable for real-time processing, whilst at the same time improving classification accuracy. To detect the connected components, contours were obtained from the binary image using the contour algorithm proposed by Suzuki et al. (1985). An oriented bounding box was fitted to each silhouette using Algorithm 1 and a selection of the bounding box metrics (height, width and hypotenuse) were measured.

To solve the problem of broken silhouettes, contours are merged based on the minimum perpendicular distances from the four corner points of each minimum fitted rectangle to the boundaries of the other bounding boxes. This was repeated for all fitted rotated bounding boxes to merge broken silhouettes (3.1). Broken silhouettes are due to noise in the video, which in some cases is resolved with the Fast Fourier Transforms (FFT).

The bounding box metrics (height, width and hypotenuse) are used to form three different 1D signals that varies with time (varying frame by frame). Mathematically, these three signals may be represented as $h(t)$, $w(t)$ and $d(t)$ respectively and $t = 0 \ldots M-1$, where M is the total number of frames in the video. Each 1D signal is then broken into short overlapping windows and the Fast Fourier Transform (FFT) (Equation 3.1) is computed for each metric (height, width and hypotenuse) separately.

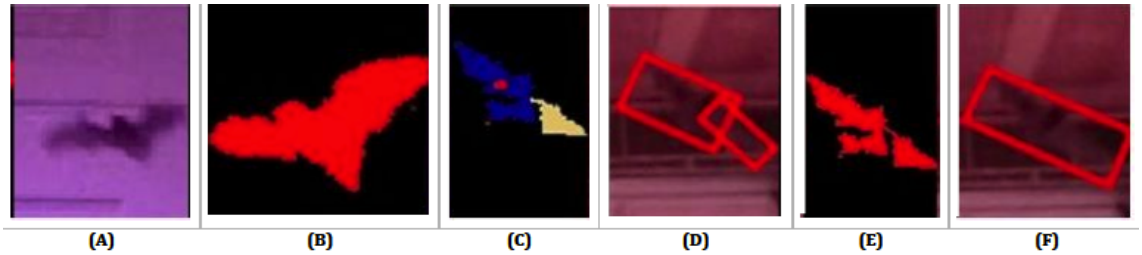


Figure 3.1: Foreground images of bats segmented using improved GMM Zivkovic and van der Heijden (2006). (A) is an original image which was segmented to achieve a perfect segmentation in (B). (C) is a case where Zivkovic and van der Heijden (2006)'s algorithm produced a broken silhouettes. There were two rotated bounding boxes fitted on the broken silhouettes (D). (E) shows when the contours were merged based on the minimum perpendicular distances from the four corner points of each minimum fitted rectangle to the boundaries of the other bounding boxes. Finally, (F) show the resultant single bounding box fitted on the bat's silhouettes.

$$F(k) = \sum_{t=0}^{(N-1)} f(t)\, e^{-i2\pi kt/N} \tag{3.1}$$

Where $f(t)$ is the signal in the spatial domain with N samples, $t = 0 \ldots N-1$ and $F(k)$ in the Frequency Domain (encoding both amplitude and phase) with $k = 0 \ldots N-1$.

$$f(t) = \frac{1}{N} \sum_{k=0}^{(N-1)} F(k)\, e^{-i2\pi kt/N} \tag{3.2}$$

Where $f(t)$ is the signal in the spatial domain with N samples, $t = 0 \ldots N-1$ and $F(k)$ in the Frequency Domain (encoding both amplitude and phase) with $k = 0 \ldots N-1$.

To determine the wing beat from the FFT stem plot, Cutler and Davis used equation 2.1. However, two techniques were proposed in this thesis, which replaces Cutler and Davis's technique.

In the first technique, the diagonal similarity technique, the dominant frequencies from each window are reconstructed into synthetic signals and used to compare with the original using the diagonal of their respective similarity matrices (3.3). The frequency which minimises this correlation is selected, and this criteria replaces that proposed by Cutler and Davis (2000). Synthetic signals were reconstructed from dominant frequencies by converting each peak in the frequency domain to a complex form (having imaginary and real parts). Signals are then formed when the inverse FFT of each peak are computed

using equation 3.2. These signals are compared with original signal and the most correlated signal's frequency is selected as the wing beat frequency. For example, figure 3.2(A) shows a sample signal taken using the hypotenuse ($h(t)$) metric with respect to time. This was transformed into the frequency domain using FFT, the stem plot of the peaks can be found in 3.2(B). For illustration purpose the first five peaks of the FFT were reconstructed into synthetic signals and their plot superimposed on the original's signal in figure 3.2(C). When the signals were correlated with the original signal, the synthetic sginal from peak 1 (blue) was found to be more correlated and therefore selected as the wingbeat of the bat. This corresponds to the ground truth wing beat frequency.

$$f_k = \sum_{i=1}^{n} |(t_{o_i} - t_{k_i})| \quad \{where\ k = 1,2,3...m\} \tag{3.3}$$

For comparison, the bounded box metrics are used to form self-similarity matrices which are used to compute the FFT. The self-similarity matrices, in this case, are computed using absolute correlation (3.4). To determine the wing beat frequency, each column of the similarity matrix is linearly de-trended and a Hanning filter applied. The result is then used to compute the power spectra for all columns of the self-similarity matrices. For accuracy, the skewness of each of the fix columns of the similarity matrices was obtained and either their spectra averaged or median estimated depending on the results of the skewness in Equation 3.5. The highest peak of $P(f_k)$ is then selected to represent the wing beat frequency.

$$S_{t_1,t_2} = \sum_{i=1}^{n} |t1_i - t2_i| \tag{3.4}$$

$$P(f_k) = \begin{cases} Mean(P(f_k)) & \text{if } |\frac{3(mean-median)}{\sigma}| < 0.5 \\ Median(P(f_k)) & \text{otherwise} \end{cases} \tag{3.5}$$

This approach uses the bounding box metrics which is 1D and is computationally efficient as opposed to the 2D images used by Cutler and Davis (2000). The diagonal similarity method in Equation 3.4 was used to select peak values even when the signal

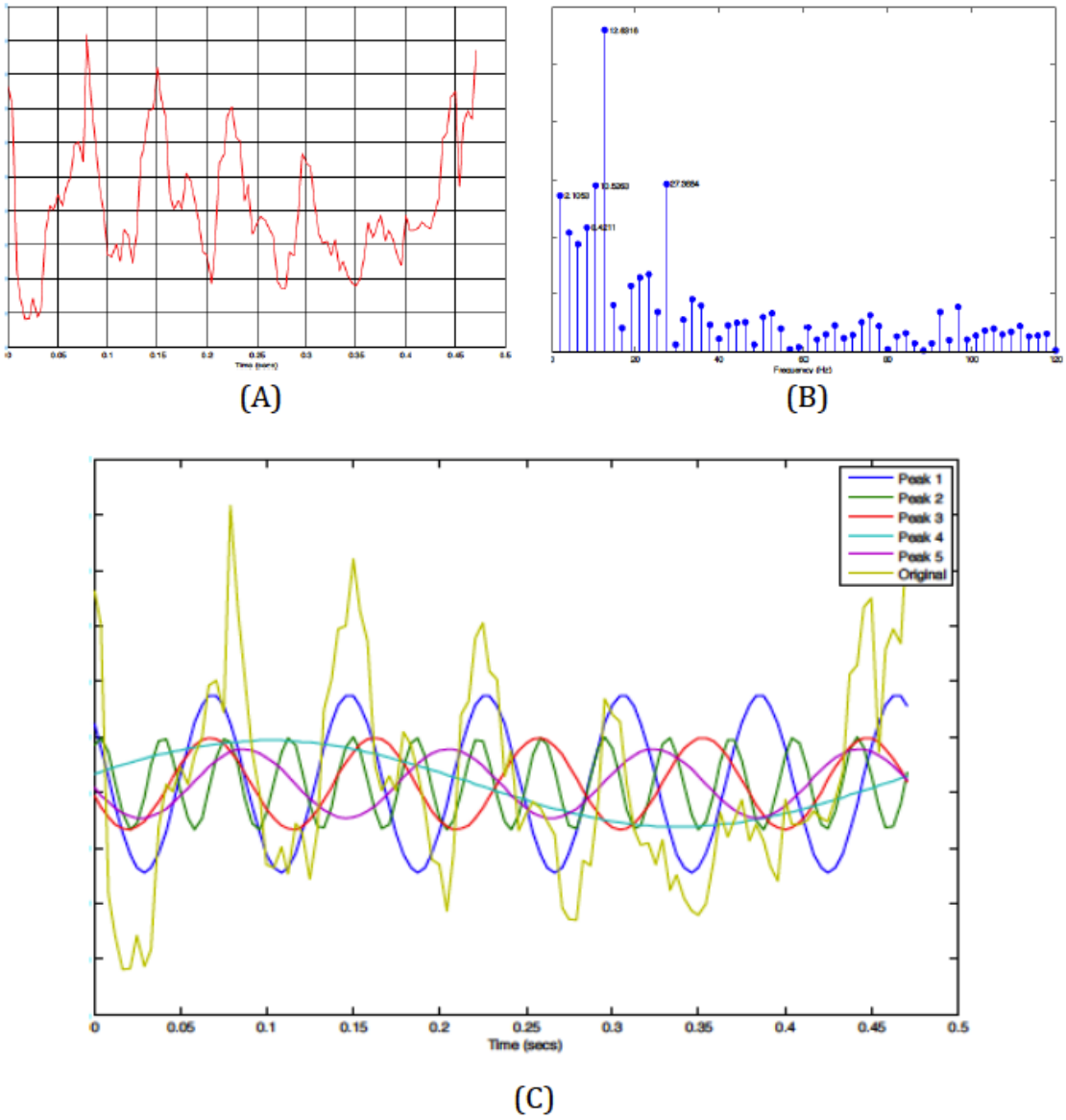
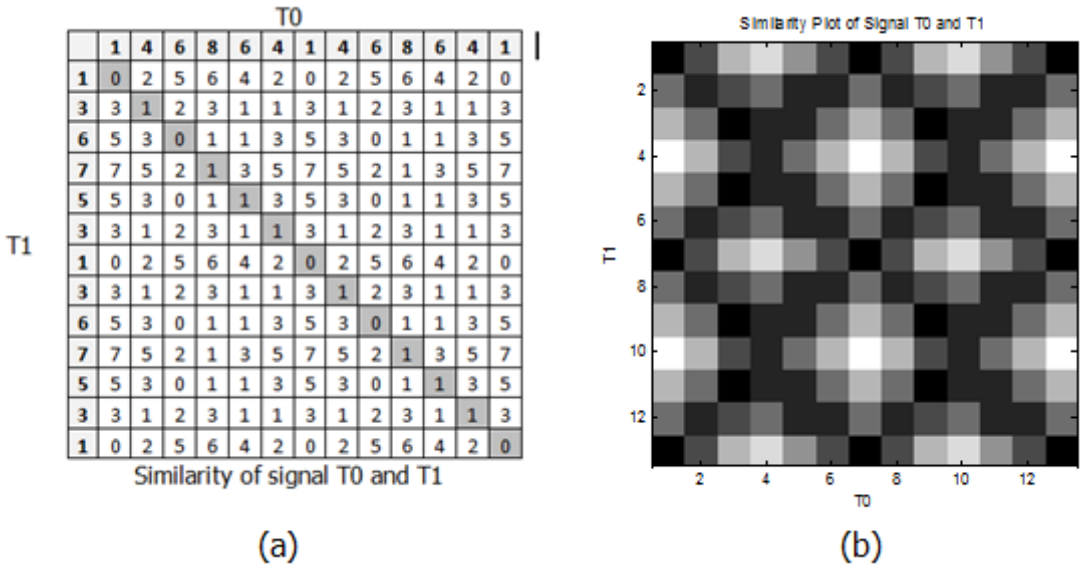


Figure 3.2: (A) Shows a sample signal taken using the hypotenuse ($h(t)$) metric with respect to time. (B) Show the FFT stem plot of the signal in (A) with each peak marked with their corresponding frequencies and (C) show the reconstructed signals for the first five peaks superimposed on the original signal (lime). The more correlated signal is peak 1, which is in fact the wing beat of the bat when compared with ground truth.

is buried in noise as opposed to Equation 2.1, which was used by Cutler and Davis to discriminate between periodic and non-periodic motion. The following section illustrates the diagonal similarity selection method using three synthetic signals.

Assuming three numerical signal: $T_0 = \{1,4,6,8,6,4,1,4,6,8,6,4,1\}$, $T_1 = \{1,3,6,7,5,3,1,3,6,7,5,3,1\}$ and $T_2 = \{1,2,4,5,4,2,1,2,4,5,4,2,1\}$. If T_0 is the

original signal then to determine which of the other two signals T_1 and T_2 is more related to T_0, the diagonal similarity matrix selection approach equation 3.4 is applied. The diagonal similarity of T_0 and T_1 is 8 (see Figure 3.3 (a)) and is smaller than that of T_0 and T_2, which is 22. The similarity plot of T_0 and T_1 can be found in Figure 3.3 (b), which has patterns that can be used to determine the peak frequency.



	1	4	6	8	6	4	1	4	6	8	6	4	1
1	0	2	5	6	4	2	0	2	5	6	4	2	0
3	3	1	2	3	1	1	3	1	2	3	1	1	3
6	5	3	0	1	1	3	5	3	0	1	1	3	5
7	7	5	2	1	3	5	7	5	2	1	3	5	7
5	5	3	0	1	1	3	5	3	0	1	1	3	5
3	3	1	2	3	1	1	3	1	2	3	1	1	3
1	0	2	5	6	4	2	0	2	5	6	4	2	0
3	3	1	2	3	1	1	3	1	2	3	1	1	3
6	5	3	0	1	1	3	5	3	0	1	1	3	5
7	7	5	2	1	3	5	7	5	2	1	3	5	7
5	5	3	0	1	1	3	5	3	0	1	1	3	5
3	3	1	2	3	1	1	3	1	2	3	1	1	3
1	0	2	5	6	4	2	0	2	5	6	4	2	0

Figure 3.3: (a) shows the similarity table of the synthetic signal and (b) shows the similarity plot resulting from the the similarity table in (a).

3.2 Experiments

Two sets of experiments were performed: the bounding box with a diagonal selection and the bounding box with self-similarity techniques. The results from these were then evaluated against the baseline algorithm proposed by Cutler and Davis (2000) and preliminary results using bats data presented. For evaluation the methods in Cutler and Davis were coded in MATLAB and since their dataset was not available the dataset introduced in this chapter was used and the results were compared. The sample videos were divided into 64 frames each with an overlap of 32 frames. For example, in sample 2, the video was divided into three 64 time frames (three windows) with 50% overlap (32 frames overlap). Therefore, window 2 of sample 2, which is made up of 64 frames, had the first 32 frames

in common with window 1. Three bats experts were used to manually count wingbeat cycles (to the nearest quarter) from the videos. The results of the ground truth can be found in Section 3.3 of this chapter. To find out if window size can improve the resolution of the wing beats, the window size was increased to 128 frames with an overlap of 64 frames and again, the three bat experts made to manually count the wingbeat cycles (to the nearest quarter). The results of this have again been reported in Section 3.3. For all these videos, the median wingbeat cycles were estimated and used to calculate the frequencies which represent ground truth data. The frequencies produced by the various algorithms that were closest to this value were then counted and the results reported.

3.3 Results

Results of experiments performed using samples described in Section 3.1.1 are presented in this section. For each window, the estimated wingbeat frequency for each method is compared to the ground truth value. The values that are the same as the ground truth are highlighted as yellow in Tables 3.1, 3.2, 3.3, 3.4 and 3.5. However, if no value is equal to the ground truth, the next closest value is selected and highlighted as yellow. Results in Tables 3.1, 3.2, 3.3 and 3.4 are to a resolution of $3.75Hz$. The resolution is calculated as $\frac{F_S}{Q}$, where F_S is the input signal's sampling rate (in our case it is 240, as the video is sampled at 240 frames per second) and $Q = 64$ is the number of FFT points used. The fundamental frequency is $0Hz$ and the other frequencies are spaced at intervals of $3.75Hz$, and are $3.75Hz, 7.50Hz, 11.25Hz, 15Hz$ and so on. Similarly, results in Table 3.5 are also to a resolution of $1.875Hz$, as $F_S = 240$ and $Q = 128$. The fundamental frequency is $0Hz$ and the other frequencies are spaced at intervals of $1.875Hz$ and are $1.875hz, 3.75Hz, 5.625, 7.5, 9.375, 11.25$ and so on.

In sample 1, there were some turning, gliding and tracking errors. The sample was divided into two short sequences and used to perform these experiments. Using the first part of sample 1, the bounding box with self-similarity matrices achieved nearly 100% closest frequencies whilst the diagonal selection and the baseline algorithm achieved 50% (see Table 3.1). The second section of sample 1 was very irregular, so the baseline algorithm

had no closest match but the two improved techniques had 50% (see Table 3.2).

Keys to all tables: (CD = Baseline Algorithm (Cutler and Davis), DS-H = Diagonal Selection - Height, DS-W = Diagonal Selection - Width, DS-HY = Diagonal Selection - Hypotenuse, SS-H = Self Similarity - Height, SS-W = Self Similarity - Width, SS-H = Self Similarity - Hypotenuse)

Table 3.1: Bat wingbeat frequencies in Hz for Sample 1 Part 1. The highlighted values represents the closest.

	C&D		**DS-H**	**DS-W**	**DS-HY**		**SS-H**	**SS-W**	**SS-HY**		**GT**
Window 0	3.75		3.75	3.75	11.25		11.25	7.25	11.25		11.25
Window 1	11.25		26.25	11.25	26.25		11.25	11.25	11.25		11.25
Closest	**1**		**0**	**1**	**1**		**2**	**1**	**2**		
%Closest	50%		0%	50%	50%		100%	50%	100%		

Table 3.2: Bat wingbeat frequencies in Hz for Sample 1 Part 2. The highlighted values represents the closest.

	C&D		**DS-H**	**DS-W**	**DS-HY**		**SS-H**	**SS-W**	**SS-HY**		**GT**
Window 0	-		3.75	3.75	26.25		3.75	3.75	15		12.19
Window 1	-		3.75	3.75	26.25		3.75	3.75	26.25		11.25
Closest	**0**		**1**	**1**	**0**		**1**	**1**	**1**		
%Closest	0%		50%	50%	0%		50%	50%	50%		

Sample 2 had more regular data with some turning and gliding made by the bat. The baseline algorithm and improved technique based on self-similarity achieved better results than the diagonal selection method. From table 3.3, the hypotenuse metric of the self-similarity metrics (SS-HY) and baseline algorithm (C&D) had 100% closest frequencies.

Table 3.3: Bat wingbeat frequencies in Hz for Sample 2. The highlighted values represents the closest.

	C&D		**DS-H**	**DS-W**	**DS-HY**		**SS-H**	**SS-W**	**SS-HY**		**GT**
Window 0	11.25		26.25	26.25	26.25		3.75	11.25	11.25		13.13
Window 1	11.25		3.75	3.75	26.25		3.75	3.75	11.25		13.13
Window 2	15		15	15	15		15	3.75	15		13.13
Closest	**3**		**1**	**1**	**1**		**1**	**1**	**3**		
%Closest	100%		33%	33%	33%		33%	33%	100%		

Sample 3, had a longer sequence, which comprises of regular and irregular data patterns with some few bat turnings and gliding. Table 3.4 shows that both the improved techniques produces better results (about 71% of the frequencies were closer to the Ground

Truth(GT)) than the baseline algorithm proposed by Cutler and Davis (36%). This is because the bat's orientation changes more rapidly in the video sequence but Cutler and Davis's techniques was not developed for this type of erratic movements in flying species. Window 10 of sample 3 is not considered periodic, so is ignored by baseline algorithm (Cutler and Davis, 2000) because there is no ideal peaks (see Figure 3.4 for window 10). This is however due to the fact that the signal is buried in noise and the baseline algorithm works strictly with only periodic motions and rejects non-periodic motions. However, this was successfully detected by the diagonal selection method and the self-similarity techniques.

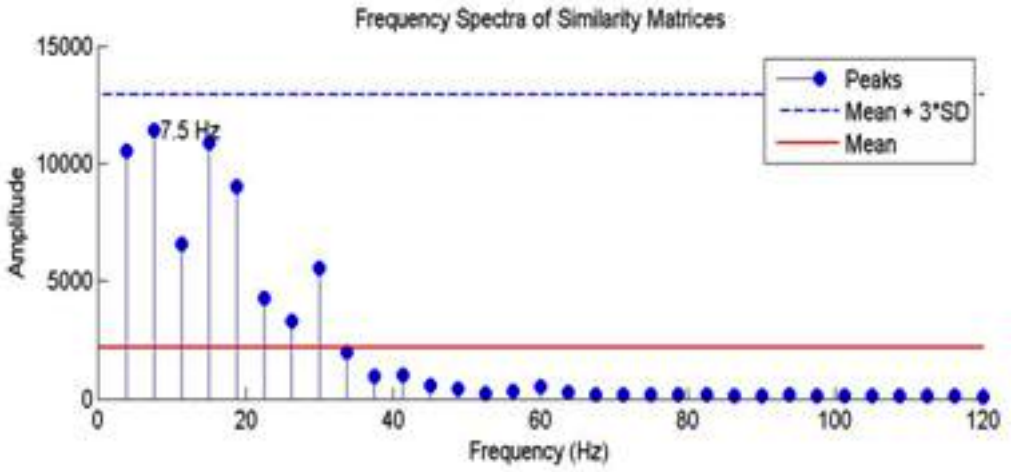


Figure 3.4: The baseline algorithm's FFT stem plot for Window 10

When only one peak lies above the threshold (Mean + 3SD), it is an ideal situation for detecting periodicity in signals. When the signal is less noisy and applying FFT completely eliminate the signal noise, then all three algorithms are able to detect the peak values correctly. For example window 3 of sample 2 comprises of regular patterns with some occasional bat turnings and gliding. Applying FFT eliminates most of the noise in the signal. In figures 3.5 and 3.8, both the self-similarity plot from the proposed method and the recurrence plot from the baseline algorithm had a regular pattern, therefore all three methods were able to filter out the correct wingbeat frequency.

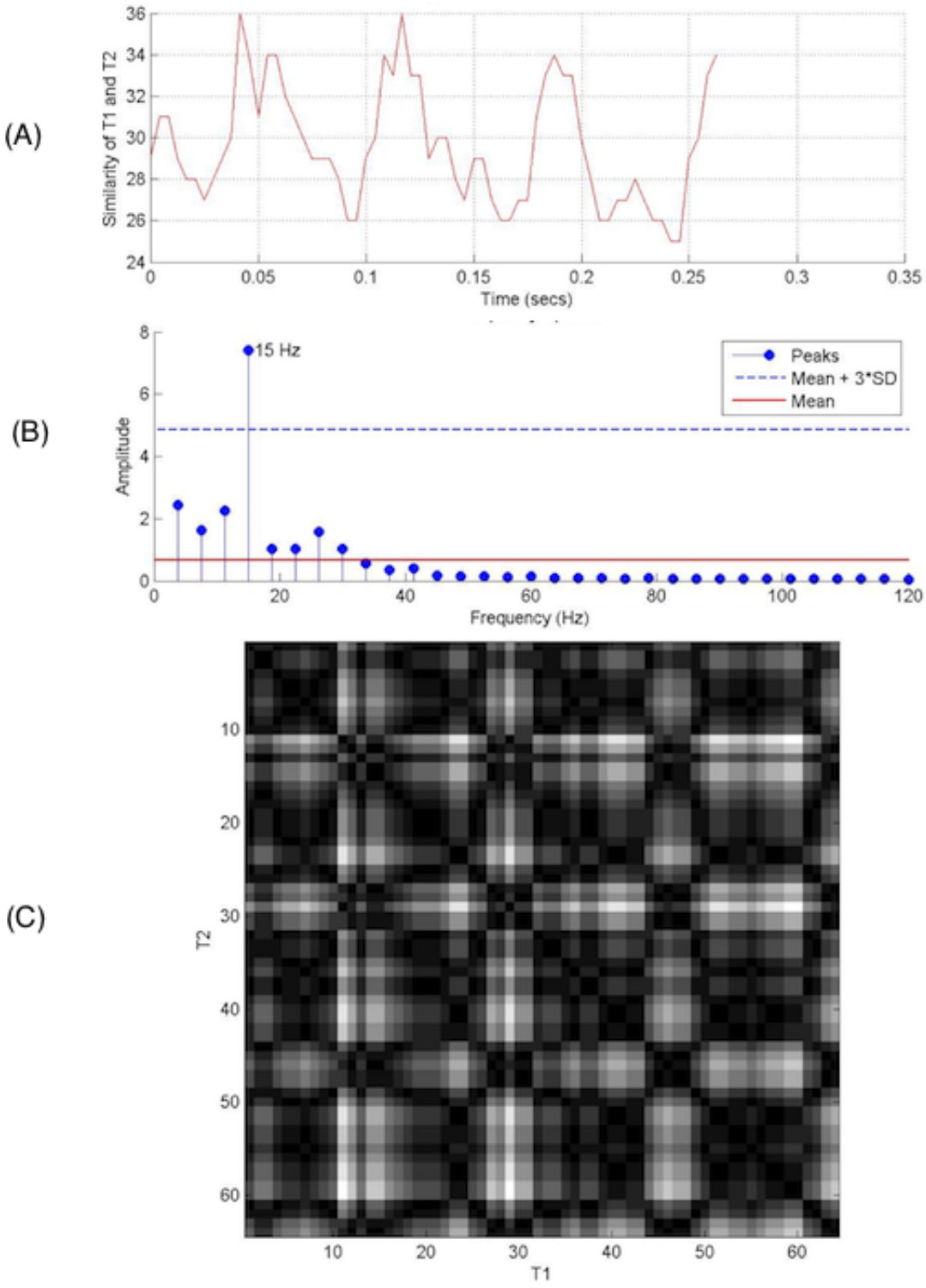


Figure 3.5: Using the proposed method, above is the siginal plot, in the middle is the stem plot and below is the similarity plot from window 2

In other cases, there were two or more peaks above the threshold (Mean + 3SD).

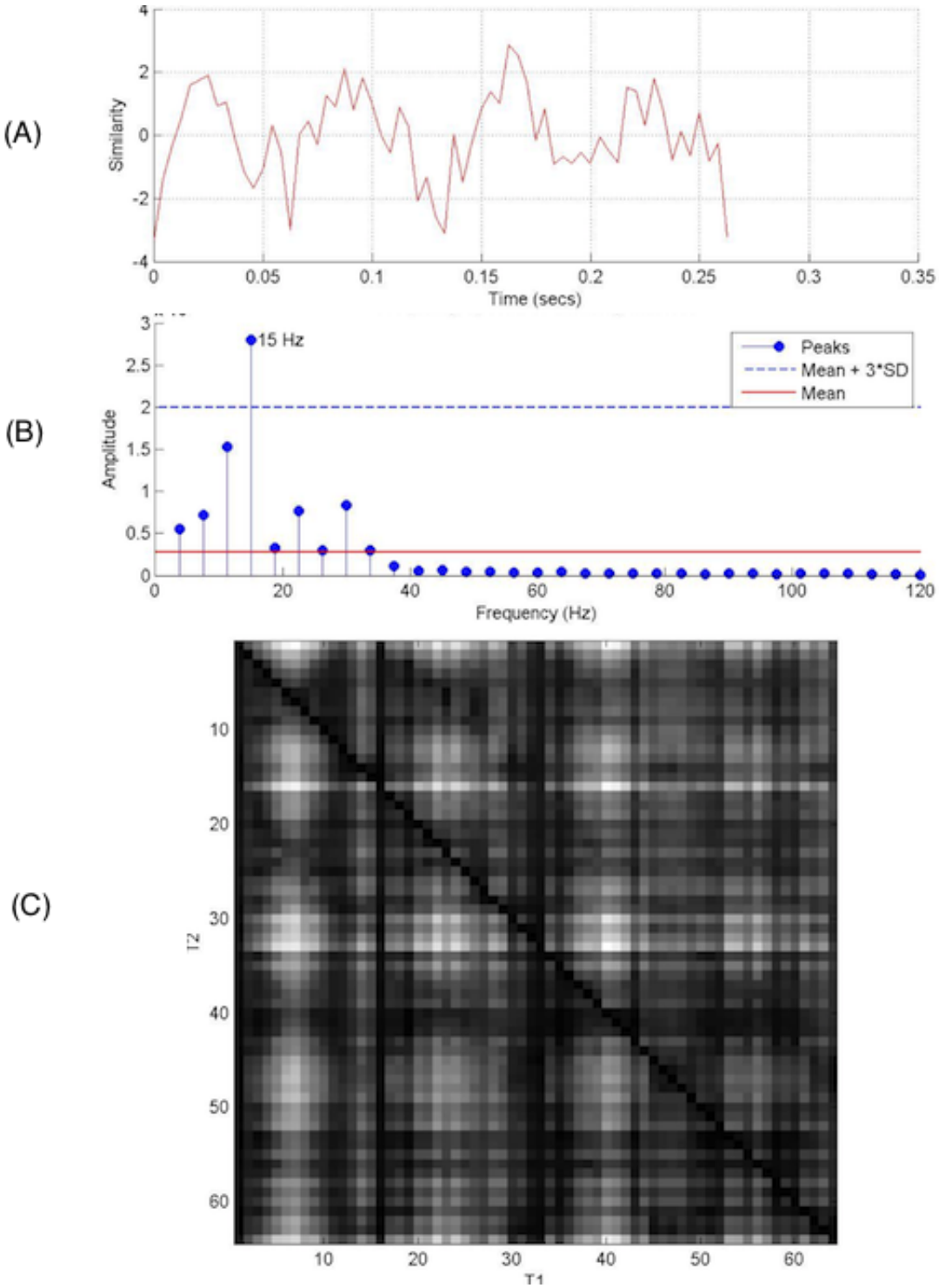


Figure 3.6: Using method by Cutler and Davis (2000), above is the siginal plot, in the middle is the stem plot and below is the similarity plot from window 2

These were not rejected by the baseline algorithm but usually, resulted in the wrong wing-beat frequency value. However, the two proposed methods were able to detect this wing beat frequency value correctly. For example window 6 of sample 3 comprises of regular and irregular patterns with occasional bat turnings and gliding. This resulted in a mixture

of clean and noisy signals and applying FFT cannot completely eliminate the noise. In figure 3.7, the self-similarity plot from the proposed method had a mixture of irregular and regular pattern. However, this method was able to filter out the correct wingbeat frequency. The recurrence plot in figure 3.8 from the baseline algorithm had irregular patterns and was not able to filter out the correct wingbeat frequency.

Table 3.4: Bat wingbeat frequencies in Hz for Sample 3. The highlighted values represents the closest.

	C&D		**DS-H**	**DS-W**	**DS-HY**		**SS-H**	**SS-W**	**SS-HY**		**GT**
Window 0	15		15	15	15		11.25	15	11.25		13.13
Window 1	15		15	15	15		11.25	15	15		14.06
Window 2	3.75		15	15	15		15	15	15		15
Window 3	3.75		15	15	15		11.25	15	15		15
Window 4	11.25		3.75	26.25	11.25		15	15	11.25		15
Window 5	3.75		11.25	11.25	11.25		11.25	11.25	11.25		9.38
Window 6	-		3.75	3.75	7.5		3.75	11.25	11.25		3.75
Window 7	15		22.5	22.5	7.5		7.5	7.5	7.5		5.63
Window 8	15		7.5	3.75	3.75		7.5	3.75	15		11.25
Window 9	11.25		3.75	3.75	15		3.75	3.75	15		12.19
Window 10	-		11.25	11.25	3.75		3.75	3.75	7.5		13.13
Window 11	3.75		11.25	11.25	33.75		11.25	11.25	15		13.13
Window 12	15		18.75	18.75	3.75		3.75	3.75	15		13.13
Window 13	3.75		52.5	52.5	26.75		11.25	11.25	15		14.06
Closest	**5**		**9**	**8**	**6**		**8**	**8**	**10**		
%Closest	36%		64%	57%	43%		57%	57%	71%		

Table 3.5: Bat wingbeat frequency in Hz for Sample 3 using 128 Window size. The highlighted values represents the closest.

	C&D		**DS-H**	**DS-W**	**DS-HY**		**SS-H**	**SS-W**	**SS-HY**		**GT**
Window 0	-		1.88	15	15		28.13	7.5	7.5		13.59
Window 1	-		1.88	13.13	13.13		9.38	7.5	54.38		12.19
Window 2	1.88		1.88	1.88	13.13		9.38	7.5	48.75		9.38
Window 3	-		1.88	1.88	1.88		9.38	7.5	7.5		6.56
Window 4	-		7.5	3.75	3.75		15	15	9.38		11.72
Window 5	1.88		11.25	11.25	1.88		16.88	11.25	9.38		15
Closest	0		0	2	2		2	1	2		
%Closest	0%		0%	40%	40%		40%	20%	40%		

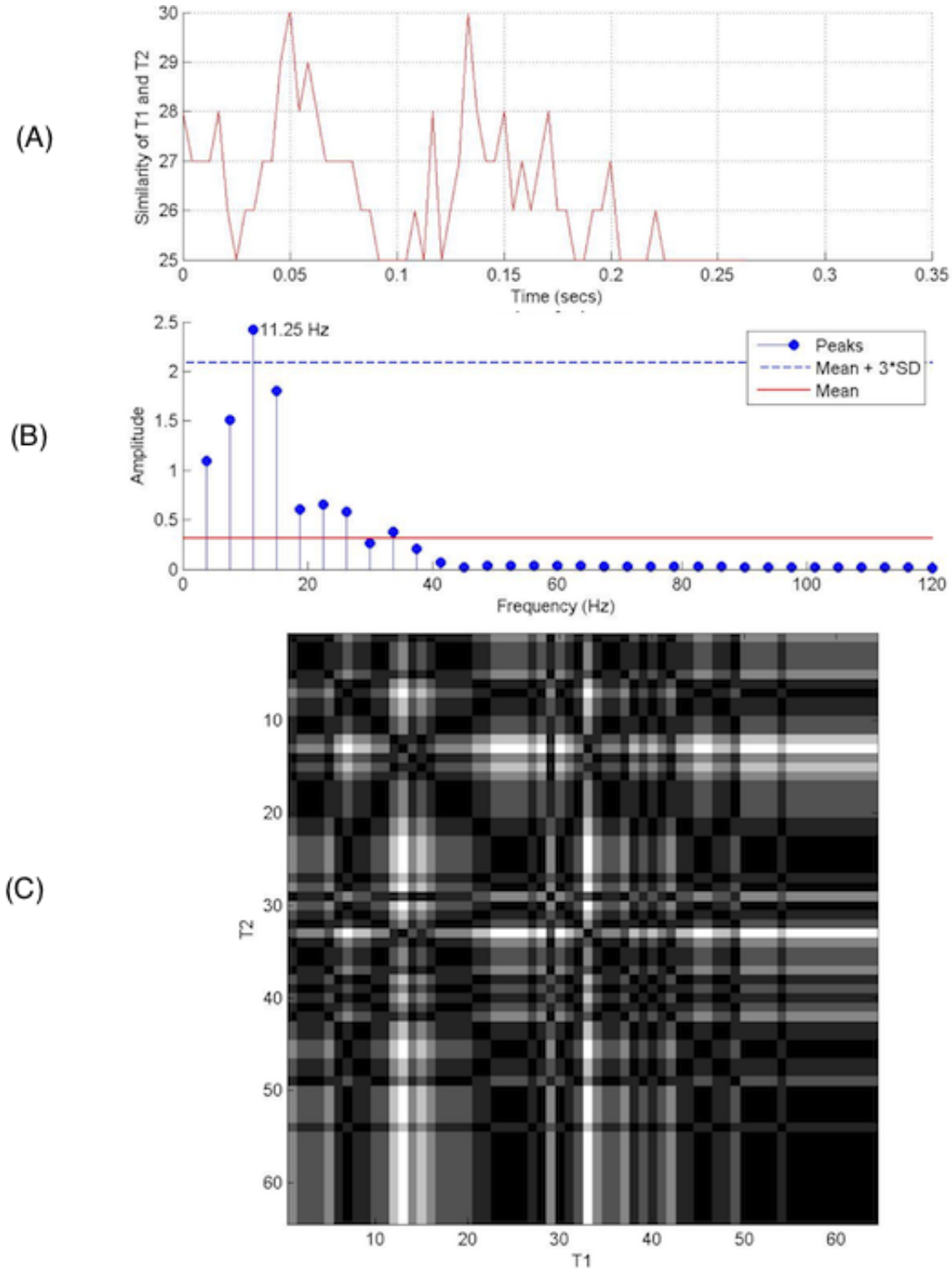


Figure 3.7: Using proposed method, above is the siginal plot, in the middle is the stem plot and below is the similarity plot from window 6

The frequency resolution of the ground truth data is finer than the results produced by the various algorithms. These algorithms produced a frequency resolution of 3.75 Hz,

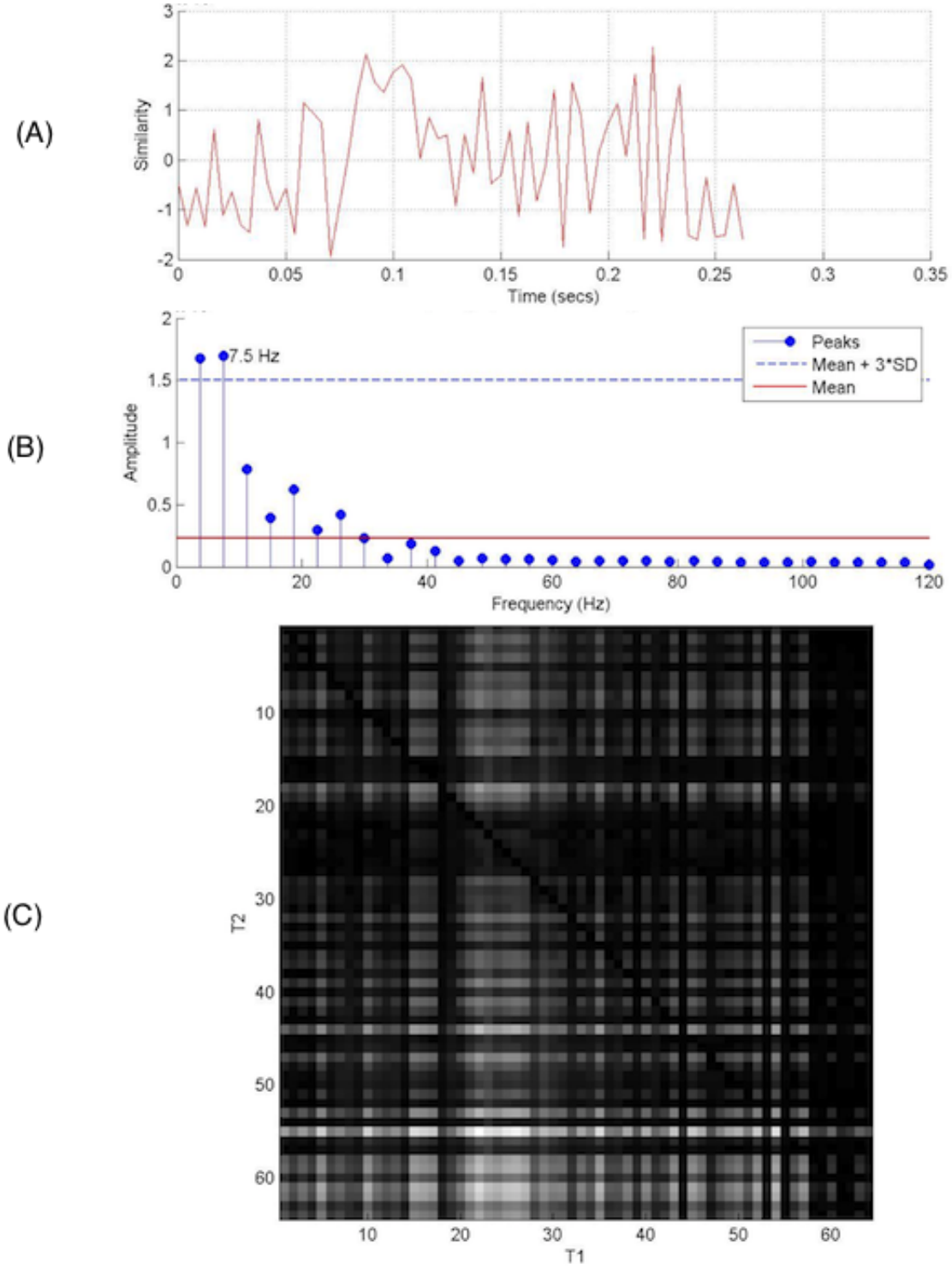


Figure 3.8: Using method by Cutler and Davis (2000), above is the siginal plot, in the middle is the stem plot and below is the similarity plot from window 6

which was coarser, compared to the ground truth resolution of 0.94Hz. Wingbeat frequencies of bats from literature for some species range from 2.99Hz to 15Hz and only vary by an average of 0.34Hz among species (Bullen and McKenzie, 2002; Norberg and

Norberg, 2012). Hence to be able to classify species, a resolution of about 0.24Hz should be achieved by the algorithms. To achieve this resolution, the 1D signal length can be increased to 512 frames or more to achieve a resolution of 0.47Hz or better. This can be done by using spectral interpolation (Lyons, 2010), increasing the frame rate to 480frames/sec or adjusting the camera to keep bats longer in the region of interest. To confirm this will work, the longer video sequence (sample 3) was divided it into 128 windows and overlap of 64. This gave a resolution of 1.88Hz and the results were presented in table 3.5. The results show a small deviation from the three algorithms compared with the ground truth data.

3.4 Conclusions

Preliminary investigation shows that both improved techniques achieved better results than the baseline line algorithm by Cutler and Davis (2000) in the estimation of wing beat frequencies when the bat's orientation changes. The self-similarity matrices with the bounding box technique used were also as good as the baseline algorithm when the bat's orientation does not change much.

When only one peak is above the threshold value on the frequency plot, all methods are able to find the correct wingbeat frequency and this situation is considered to be ideal by the baseline algorithm Cutler and Davis (2000). When there are more than one peak values above the threshold, the proposed methods are able to find the wingbeats that closely match the ground truth compared with the baseline algorithm. Finally, when all peaks are below the threshold, the proposed methods are able to estimate more wingbeats frequencies which are closer to the ground truth than the baseline algorithm.

Increasing the window size can help improve the frequency resolution, thus improving the accuracy of the wingbeat frequencies. Two ways of improving the frequency resolution are increasing the time domain sample and or using spectral interpolation Lyons (2010). The fast Fourier transform (FFT) was used in Chapter 5 to extract wingbeat frequencies features for the bird species classification algorithms.

The following chapter will explore the classification of bird species using appearance-based features and will propose an extended set of appearance features which will be evaluated with the feature set proposed by Marini et al. (2013).

Chapter 4

Classification of Bird Species using Appearance Features

A small but growing number of studies (Marini et al., 2013; Wah et al., 2011a; Duan et al., 2012; Berg and Belhumeur, 2013; Branson et al., 2014; Welinder et al., 2010), have used computer vision with image-based (appearance) techniques, which potentially provide richer and more informative data (continuous position, behaviour, and other physical features), to identify and classify species. The fine-grained appearances of some bird species have made their classification a very challenging task especially for methods that uses colour and or shape alone (Marini et al., 2013; Welinder et al., 2010). For this reason, part-based modelling (Berg and Belhumeur, 2013; Branson et al., 2014; Duan et al., 2012; Berg et al., 2014), which requires the birds' parts to be manually marked prior to automation are also emerging.

Birds are typically categorized by their shape or silhouettes and physical characteristics. Visual properties (e.g., shape, colour etc.) are important keys for bird recognition. Often you don't need to see any colour at all to know what kind of bird you're looking at. The work in this section was motivated by the fact that silhouettes provides additional characteristics such as size, proportions, and posture to help differentiate species that are closely rated in colour. This chapter introduces initial experiments performed to classify birds in flight from sequences of still images using the extended appearance features (colour and shape), with standard classifiers such as Naive Bayes (NB), Random Forest

(RF), Random Trees (RT) and Support Vector Machine (SVM). The aim is to identify a set of appearance features used in previous related works and refine them for use in automatic classification of species in flight. Three datasets are used for evaluation, which includes a video dataset of seven species, an extended videos dataset of thirteen classes (eleven bird species, one with three colour forms) and Caltech-UCSD BIRDS-200-2011. The identified feature sets were compared with existing state-of-the-art image-based method (Marini et al., 2013) and the results presented in this chapter. The rest of the chapter is structured as follows:

- Section 4.1 presents the materials and methods used for the appearance features classification including overview of the processing methods and conclude with details of statistical features used for the classification models.
- Section 4.2 describes the appearance feature set used for all experiments in this chapter.
- The experimental setup and results are then described in Sections 4.3 and 4.4 respectively.
- Finally, Section 4.5 presents conclusions, which include summarising all results in the chapter and introducing briefly what will be described in the following chapter.

4.1 Datasets and Methods

This section introduces the three datasets used to perform experiments throughout this thesis. It also presents discussion of methods, including pre-processing, feature extraction and the statistical features computed for classification of species.

4.1.1 Datasets

Three datasets are used for all experiments in this chapter, of which two were collated purposely for this thesis and future research. These to the best of our knowledge are the only existing video-based birds dataset, giving the opportunity to perform classification using motion features. The first which is known as "dataset #1", is a set of videos covering

seven classes, made up of seven different bird species in flight. This has a total of 170 videos, representing 23,932 image sequences. The second referred to as "dataset #2", is an extended set of videos covering thirteen classes made up of eleven bird species, one (Budgerigar (*Melopsittacus undulatus*)) with three colour forms, also in flight. This has a total of 957 videos, representing 161,907 image sequences. These datasets were collated because there is currently no bird species datasets which incorporate motion. The last, "Caltech-UCSD Birds-200-2011" dataset, has been extensively applied to fine-grained classification using computer vision. This is a still image data set of birds compiled by Wah et al. (2011b) from the internet, which contains 200 species, with a total of 11,788 images.

Figure 4.1: Segmented birds from the dataset #1 using the method in (Zivkovic and van der Heijden, 2006). From left to right: Common Wood Pigeon, Superb starling, Nanday Parakeet, Budgerigar (wild-type) and Cockatiel.

Dataset #1 and #2 were recorded using a Casio Exilim ZR100, mounted on a tripod and recording at 240 frames per second. The videos were recorded over different days over two years and from various sites, which includes the National Parrot Zoo, and my backyard garden in Lincoln. Most of the species recorded consist of more than 15 individuals, apart from Superb Starlings, which had only three. Dataset #1 was the first collated and consist of seven bird species, which is made up of 170 short videos (representing 23,932 image frames) of approximately two to five seconds. Details of this has been provided in Table 4.1 and samples in Figure 4.1. This dataset has been used to perform experiments in Chapters 4 - 7. The majority class (class with the most samples) for these samples is the Common House Martin (*Delichon urbica*), which is made up of 30 videos and the minority (class with the least samples) is the Common Wood Pigeon (*Columba palumbus*), which is made up of 17. The dataset #2 consists of thirteen classes (eleven

bird species, one with three colour forms), and is an extension to dataset #1 with seven new classes but without the superb starling class. The superb starling class was omitted from dataset #2 as there were limited number of samples (videos) for this species (21 videos) as opposed to the minimum of 35 videos used for each class of this dataset. This consists of 957 short videos with 161,907 images. This has more classes and samples than dataset #1 and was collated to help understand the effects of increasing the classes in the datasets on classification rate and to make the dataset more challenging for future studies (more related species).

Again, this has been used to perform experiments in Chapters 4 - 7. Details and samples of this dataset have been provided in Table 4.2 and Figure 4.2 respectively. Most of the classes in this dataset have more than 37 videos. The minority class has 37 videos, whiles the majority has 147.

There are three species that have very closely related appearances in dataset #2, these are referred to as "species with fine-grained appearances". They are the Alexandrine Parakeet (*Psittacula eupatria*), Nanday Parakeet (*Aratinga nenday*) and Blue-crowned Parakeet (*Thectocercus acuticaudatus*). One particular species, the Budgerigar, has three different colour forms, which were used to form three different classes (this is to make the motion part of the system more challenging). Table 4.3 show all the thirteen classes with their distinguishing features. All the datasets were slightly imbalanced, which means that there is a class with fewer samples, known as the minority and another with the majority samples, known as the majority class. The imbalanced ratio for dataset #1 is approximately 1.76 and that of dataset #2 is 3.97. The imbalanced ratio is defined as the ratio of the number of samples in the majority class to that of the minority (García et al., 2012; Orriols-Puig and Bernadó-Mansilla, 2009). The two dataset are highly challenging and can be used in research involving motion, species with fine-grained appearances and imbalanced datasets. However, it's more suitable for classification of species in flight from videos and for trajectories categorisation.

Table 4.1: Table showing the number of videos and images in the seven species dataset taken at three different sites, with majority from The National Parrot Sanctuary, Lincolnshire, UK

Species	# of videos	# of images
Common House Martin	30	4,378
Common Wood Pigeon	17	4,347
Superb Starling	21	1,920
Nanday Parakeet	27	4,155
Cockatiel	25	3,942
Common Starling	23	2,914
Budgerigar (wild-type)	27	2,276
Total	**170**	**23,932**

The last dataset, Caltech-UCSD Birds-200-2011 is an extended version of the CUB-200 dataset, with roughly double the number of images per class. All images have been annotated with bounding boxes, part locations, and attribute labels. The images and annotations have been filtered by multiple users of Mechanical Turk. There are 11,788 still images in this dataset covering 200 categories (species). Each category contains 60 or more images with 15 part locations per image, and a bounding box. Figure 4.3 shows images of some species taken from the Caltech-UCSD Birds-200-2011 dataset. This dataset was chosen because it has been well validated and is one of the few bird species datasets which has been well used in computer vision for still image classification; and it has also been used by Marini et al. (2013), the work which is used in the evaluation of the proposed appearance features.

Table 4.2: Table showing the number of videos and images in thirteen classes (eleven bird species, one with three colour forms) dataset taken at three different sites, with majority from The National Parrot Sanctuary, Lincolnshire, UK

Species	# of videos	# of images
Alexandrine Parakeet	79	12,801
Nanday Parakeet	60	10,025
Blue-crowned Parakeet	60	9,076
Common House Martin	139	25,517
Eastern Rosella	44	5,929
Budgerigar (yellow)	54	7,667
House Sparrow	78	10,191
Budgerigar (wild-type)	48	6,283
Common Wood Pigeon	37	4,301
Black-headed Gull	147	38,764
Cockatiel	59	9,398
Budgerigar (blue)	81	12,090
Common Starling	71	9,865
Total	**957**	**161,907**

4.1.2 Sample Size Per Category

The number of samples per categories that will improve classification rate have been investigated. This helped us to make an informed decision as to the number of samples needed for each species during data collection. These experiments were performed with the initial dataset #1 and it was observed that from a sample of 200 to 700 per category, the classification rate increases sharply. Beyond 700 samples the improvement is less significant, but peaks at approximately 2000 samples per category (Figure 4.4).

Table 4.3: Species used in this thesis, with their distinguishing features.

Species	Distinguishing Features
Alexandrine Parakeet	Green parakeet with a patch of brilliant red on the shoulder. It has a long green tail. The male has pink collar on the back of the neck. It is about 58cm long.
Blue-crowned Parakeet	Green parakeet with a blue head and red-orange highlights in long tail. It ha a white eye ring with pink and black bill. It is about 37cm long.
Nanday Parakeet	Green parakeet with black facial mask and beak.It shows black trailing flight feathers on its wings. It is between 32-37cm long.
Common House Martin	It has metallic dark blue mantle and crown. Its wings and tail are black. It has a forked tail with solid white rump that distinguishes it from other swallows. It is about 13cm long.
House Sparrow	It has black-streaked brown upper parts with a pale gray underparts. It wings are brown with a white bar. It has black conical bill, throat and upper breast. The females lack the black colour. It is between 14-18cm long.
Common Starling	It is mainly black with buff edged wing feathers and reddish-brown legs. It is about 22cm long.
Budgerigar (wild-type)	It has a light green body colour with yellow face. Their back and wing display pitch-black mantle markings. It is about 18cm Long.
Budgerigar (yellow)	It has a yellow body colour. Their back and wing display pitch-black mantle markings. I is about 18cm Long.
Budgerigar (blue)	It has a light blue body coloure. Their back and wing display pitch-black mantle markings. It is about 18cm Long.
Common Wood Pigeon	It is grey with a white neck and wing patches, whih is visible whiles in flight. It is between 38-44cm long.
Black-headed Gull	It is a white gull with a partial hood. Its back is light-gray. it has Red bill. it is about 34-37cm long.
Cockatiel	It is grey with prominent white flashes on the outer edges of wings. The male has white or yellow face whilst the female has grey or light grey. it is between 30 - 33cm long.
Eastern Rosella	Colours: Red, white, yellow, fading green, black and grey. It has a red head, neck and breast with distinctive white cheek patches. It is 30cm long.

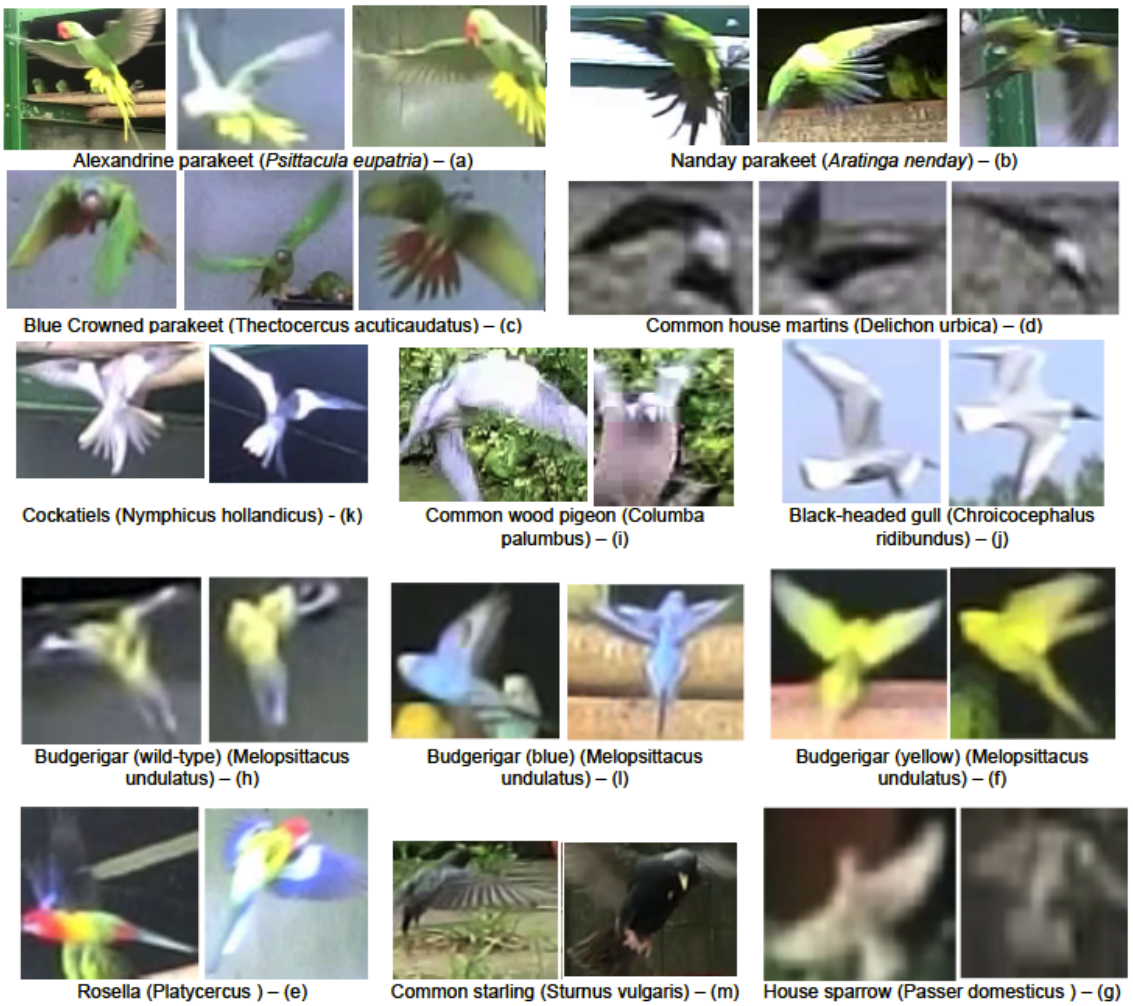


Figure 4.2: Samples from the dataset of eleven species, one of which is of different colour forms. The samples are made up of: **a)** Alexandrine Parakeet, **b)** Nanday Parakeet, **c)** Blue-crowned Parakeet, **d)** Common House Martin, **e)** Eastern Rosella, **f)** Budgerigar (yellow), and **g)** House Sparrow, **h)** Budgerigar (wild-type), **i)** Common Wood Pigeon, **j)** Black-headed Gull, **k)** Cockatiel, **l)** Budgerigar (blue), and **m)** Common Starling

4.1.3 Methods

For each video, the birds silhouettes were extracted, samples of the extracted silhouettes provided in Figures 4.1 and 4.5 for dataset #1 and #2 respectively. This was done using the background Gaussian mixture model proposed by Zivkovic and van der Heijden (2006) and which have been introduced and used in Chapter 3. To detect the connected components, contours were obtained from the binary image using the contour algorithm proposed by Suzuki et al. (1985), also introduced in the previous chapter. An oriented bounding box was fitted to each silhouette and a selection of metrics (height, width and

Figure 4.3: Sample images taken from the Caltech-UCSD Birds-200-2011 dataset Wah et al. (2011b)

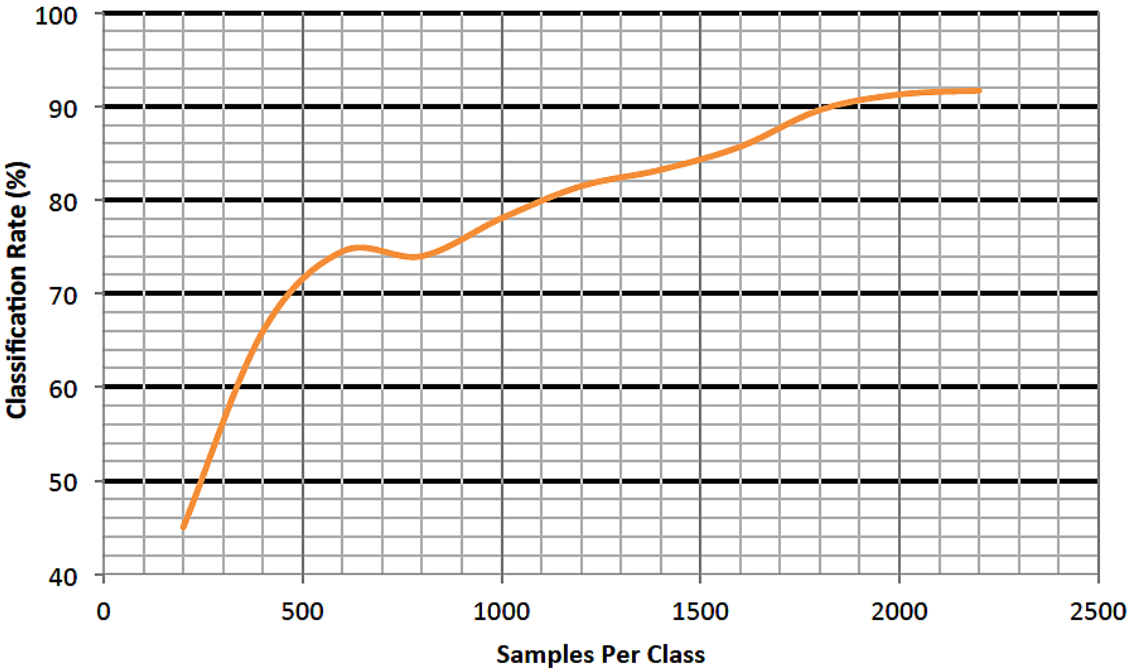


Figure 4.4: Sample size per category against classification rate (%)

hypotenuse, centroid, silhouette and contour points) were measured. Colour moments, shape moments, greyscale histogram, Gabor filter and log-polar features were extracted

from the segmented silhouettes, which have been described in detail in Section 4.2. These features were concatenated to form a feature vector for classification of bird species by colour, shape and texture.

In most cases, features were represented as statistical features, which provide information on the location, variability and appearance of the distribution of data. The statistical features computed include the mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings. The details of these features have been provided in the next section.

Figure 4.5: Segmented birds from dataset #2 using the method in (Zivkovic and van der Heijden, 2006). From left to right: Alexandrine Parakeet, Cockatiel, Nanday Parakeet, Wood pigeon, Eastern Rosella, Cockatiel #2, and Wood Pigeon #2.

4.1.4 The Statistical Features

Almost all features extracted were represented as computed statistical features, which were used to form a feature vector for classification of species. This was done to ensure that the performance of the classification model is faster, as larger feature sets impact on the processing time and may also affect the classification rate negatively (Yu and Liu, 2003). For clarity, these statistical features have been introduced briefly in the following sections.

The First-order Histogram

The first-order histogram probability provides information about the distribution of the intensity level of an image (Sergyan, 2008) and can be used to represent an image. Given an image I of size N by M; the first-order probability is defined as:

$$P_i = \frac{n_i}{NM} \tag{4.1}$$

Where i is the intensity value at a point in the image, P_i is a probability distribution of intensity value i and n_i is the number of intensity value i in the image I; the value of $0 \leq P_i \leq 1$ and $\sum_{i=0}^{NM} P_i = 1$. From the probability distribution function, other statistical features are extracted as introduced in the following sections.

The Mean of the First-order Histogram

The mean of the first-order histogram probability describes the general brightness of the image. A high mean implies a bright image and a low mean implies a dark image. The mean is defined as:

$$\bar{p} = \sum_{i=0}^{k} i(p_i) \tag{4.2}$$

Where $k = 256$ for saturation and value and $k = 180$ for hue colour channels (channel values are normalised). Where a grayscale image is used, $k = 256$.

The standard Deviation of the First-order Histogram

The standard deviation of the first-order histogram probability describes the contrast of the image. A high variance implies a high contrast and a low variance implies low contrast. The standard deviation is defined as:

$$\sigma = \sqrt{\sum_{i=0}^{k} (i - \bar{p})^2 * p_i} \tag{4.3}$$

Where $k = 256$ for saturation and value and $k = 180$ for hue (channel values are normalised). $\bar{p}$ is the mean of the first-order histogram probability. Where a grayscale image is used, $k = 256$.

The Skewness of the First-order Histogram

The skewness of the first-order histogram Probability measures the asymmetry about the mean in the intensity level distribution. The skewness is defined as:

$$skewness = \frac{1}{\sigma^3} \sum_{i=0}^{k} (i - \bar{p})^3 * p_i \tag{4.4}$$

Where $k = 256$ for saturation and value, $k = 180$ for hue (channel values are normalised) and $\sigma \neq 0$. $\bar{p}$ is the mean of the first-order histogram probability. Where a grayscale image is used, $k = 256$.

The Energy of the First-order Histogram

The energy describes the intensity levels in the image. A high energy tells us that the number of intensity levels in the image is few. This means that the distribution is concentrated in only a small number of different intensity levels. The energy is defined as:

$$energy = \sum_{i=0}^{k} |p_i|^2 \tag{4.5}$$

Where $k = 256$ for saturation and value, $k = 180$ for hue (channel values are normalised) and $\bar{p}$ is the mean of the first-order histogram probability. Where a grayscale image is used, $k = 256$.

The Entropy of the First-order Histogram

The entropy of the first-order histogram probability measures how many bits are need to code the image data. The entropy increase as the pixel values in the image are distributed among more intensity levels. This measure is inversely proportional to the energy levels. The entropy is defined as:

$$entropy = -\sum_{i=0}^{k} (p_i * (log(p_i)/log(2))) \tag{4.6}$$

Where $k = 256$ for saturation and value, $k = 180$ for hue (channel values are normalised) and $\bar{p}$ is the mean of the first-order histogram probability. Where a grayscale image is used, $k = 256$.

The Kurtosis of the First-order Histogram

The Kurtosis is used to measure the peak of the distribution of the intensity values around the mean (Malik and Baharudin, 2013). A high kurtosis datasets usually have a peak near the mean and this declines rather rapidly with heavy tails whiles those with low kurtosis have a flat top near the mean. In some cases you will have a uniform distribution which

is an extreme case. The Kurtosis can be computed as:

$$Kurtosis = \frac{1}{\sigma^4} \sum_{i=0}^{k} (i - \bar{p})^4 * p_i \tag{4.7}$$

Where $k = 256$ for saturation and value, $k = 180$ for hue (channel values are normalised) and σ is the standard deviation of the first-order histogram probability and $\sigma \neq 0$. $\bar{p}$ is the mean of the first-order histogram probability. Where a grayscale image is used, $k = 256$.

The Local Maxima and Local Minima of the First-order Histogram

Local maxima and minima occur at critical points where the derivative of the first-order probability function is zero. These are usually peaks and valleys in the distribution and provides information on a set of dominant and less dominant intensity values. In an image, peaks will represent areas of high-intensity and valleys, low-intensity and may be relevant features because they mark important image objects. In the first-order probability function, there may be multiple regional maxima and/or minima but there can only be a single global maxima or minima. These features have been used, by counting the number of minima and maxima in the first-order function, which is used to statistically represent part of the feature sets. Algorithm 2 and 3 are used to count the minima and maxima of the first-order histogram function respectively.

Algorithm 2: Find the number of local minima in the first-order histogram

```
1 LocalMin ← 0;
2 vector < float > probs;
3 for i = 0 to probs.size() − 2 do
4 |   if probs(i) − probs(i+1) ≥ 0 and probs(i+1) − probs(i+2) ≤ 0 then
5 |   |   LocalMin ← LocalMin++
6 |   end
7 end
8 Return LocalMin
```

The Maximum and Minimum of the First-order Histogram

The maximum and minimum provide information on the dominant and less dominant intensity values of the image. To compute the minimum and maximum, the probability

Algorithm 3: Find the number of local maxima in the first-order histogram

```
LocalMax ← 0;
vector < float > probs;
for i = 0 to probs.size() − 2 do
    if probs(i) − probs(i+1) ≤ 0 and probs(i+1) − probs(i+2) ≥ 0 then
        LocalMax ← LocalMax++
    end
end
Return LocalMax
```

distribution of intensity values are passed to Algorithms 4 and 5 respectively.

Algorithm 4: Find the Minimum intensity from the first-order histogram

```
Min ← 0;
MinIntensity ← 0;
vector < float > probs;
for i = 0 to probs.size() − 1 do
    if probs(i) ≤ Min then
        Min ← probs(i) MinIntensity ← i
    end
end
Return MinIntensity
```

Algorithm 5: Find the Maximum intensity from the first-order histogram

```
Max ← 0;
MaxIntensity ← 0;
vector < float > probs;
for i = 0 to probs.size() − 1 do
    if probs(i) ≥ Max then
        Max ← probs(i) MaxIntensity ← i
    end
end
Return MaxIntensity
```

The Zero Crossings of the First-order Histogram

This is the rate of sign-changes along the first-order histogram probability distribution. Gouyon et al. (2000) have used zero crossing as a key feature to successfully classify percussive sounds. In this thesis, the number of zero crossing in the first-order probability distribution function were counted (see Algorithm 6), which have been used to represent some of the features.

Algorithm 6: Find the number of zero crossing in the first-order histogram

```
probsMean ← Mean(probs);
for i = 0 to probs.size() − 1 do
    probs(i) ← probs(i) − probsMean;
end

for i = 0 to probs.size() − 2 do
    Sign1 ← Sign of probs(i) Sign2 ← Sign of probs(i + 1)

    if Sign1 ≠ Sign2 then
        zerCross ← zerCross + +;

    end

end

Return zerCross
```

4.2 Appearance Features Extracted

The appearance features introduced in this section have been developed from existing works on species classification but uses a combination of colour, shape and texture features. Specifically, features introduced and used in this thesis include colour moments and colour log-polar values, shape moments, Gabor filters and greyscale histograms.

4.2.1 Colour Moments Features

A popular way of describing colour images is using colour histogram which has been detailed in section 2.3.1 of chapter 2. They have been used as part of our appearance features by transforming colour images from RGB to HSV colour space before constructing a colour histogram (Figures 4.6 and 4.7), from the silhouette pixels, which excludes background pixels. The histogram is then built for each colour channel separately and the first-order histogram probability (Sergyan, 2008) computed. 30 bins were used for the Hue channel histogram and 32 each for the Saturation and Value channels. The frequencies of the bins were computed to form three different histograms for each of the channels. A set of five statistical features including the Mean, Standard Deviation, Skewness, Energy and Entropy (see Section 4.1.4) were then calculated from each of the histograms separately. Each channel's frequencies were concatenated with their computed statistics to form the channel's feature set. The channels features were then concatenated to form

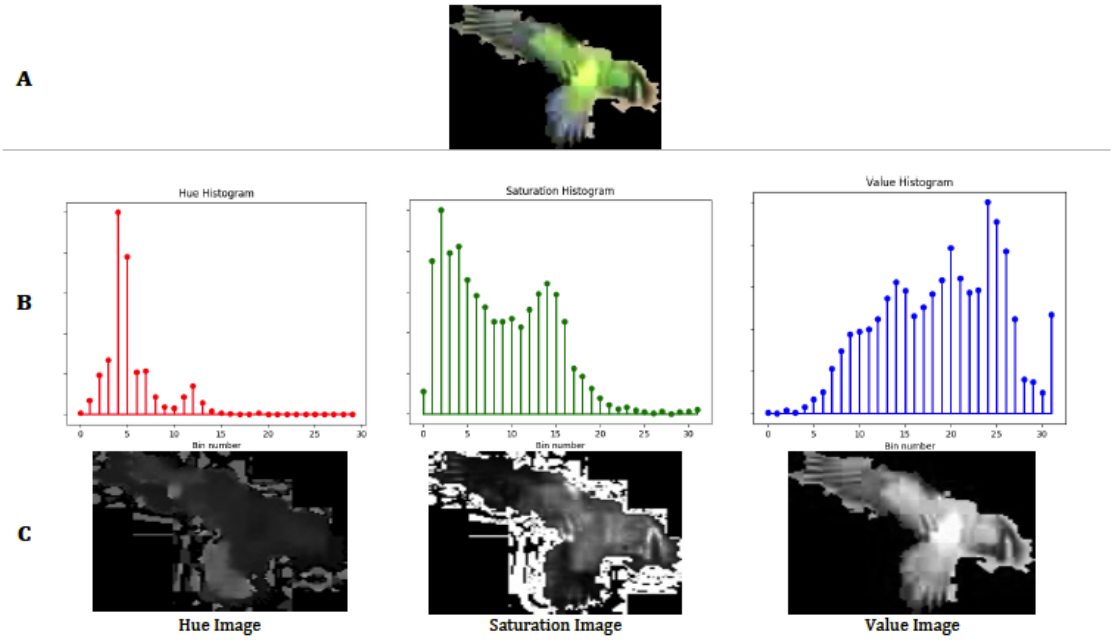


Figure 4.6: A) The original image of a Nanday Parakeet in RGB. B) The middle images in this figure show the Hue histogram (represented as red stem plot), saturation histogram (represented as green stem plot) and value histogram (represented as blue stem plot) respectively. The histograms were created using only the silhouette of the bird and without the background. The horizontal axis shows the bin numbers and the vertical, number of intensity values. For example, saturation has 32 bins, the first bin representing intensity values from 0-7, the second from 8-15 etc. C) The transformed original image separated into three images representing the three HSV channels (Hue, Saturate and Value respectively). These have been extracted separately and used to form histograms features for classification.

the colour features. In total there were 109 colour features, which is made up of 35 Hue, 37 Saturation and 37 Value features.

4.2.2 Shape Moments Features

Another important feature apart from colour for identifying birds is it's shape. To describe the shape of an object various image moments can be extracted from the image contours (Du et al., 2007). An image moment is a weighted average of the image pixels' intensities. Two moments were used in the extraction of shape features for bird species classification, which include spatial (raw) moments (Jacob et al., 2001) and Hu moments (Du et al., 2007), which have been detailed in section 2.3.1 of chapter 2. Hu moments are invariant to some transformations, such as rotation, scaling, and translation (Martín et al., 2010) and are therefore well suited for flying bird species classification.

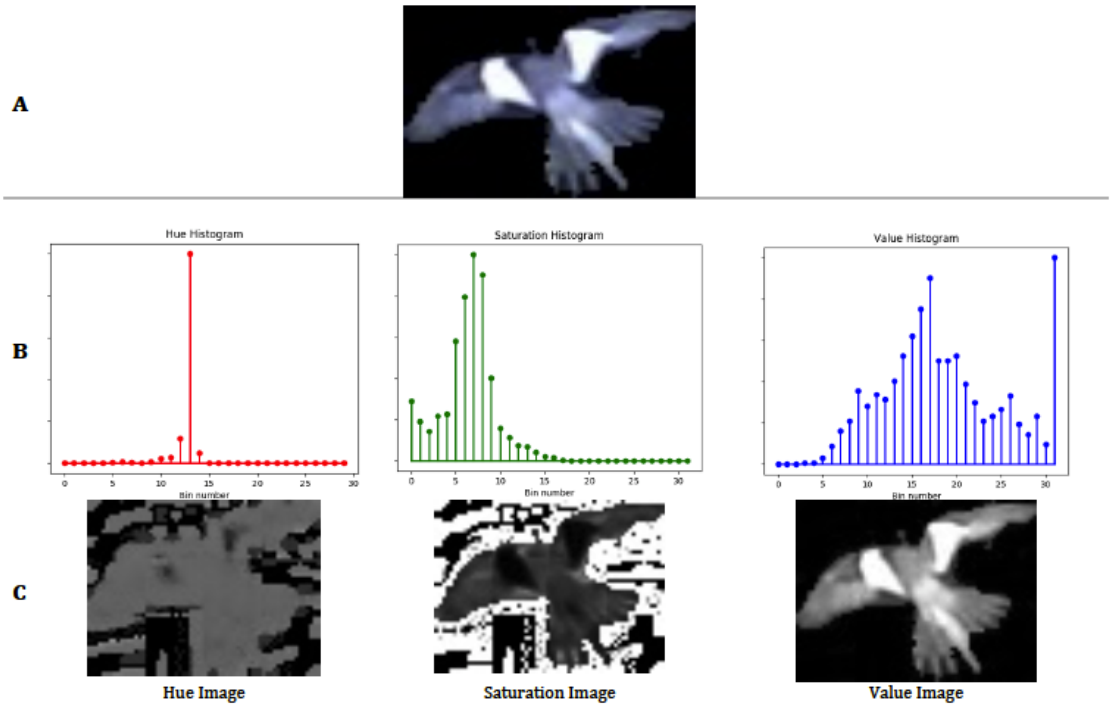


Figure 4.7: A) The original image of a Cockatiel in RGB. B) The middle images in this figure show the Hue histogram (represented as red stem plot), saturation histogram (represented as green stem plot) and value histogram (represented as blue stem plot) respectively. The histograms were created using only the silhouette of the bird and without the background. The horizontal axis shows the bin numbers and the vertical, number of intensity values. For example, saturation has 32 bins, the first bin representing intensity values from 0-7, the second from 8-15 etc. C) The transformed original image separated into three images representing the three HSV channels (Hue, Saturate and Value respectively). These have been extracted separately and used to form histograms features for classification.

To represent shape information, the shape contours (Suzuki et al., 1985) were extracted and seven Hu moments and ten spatial moments were then extracted from them (see Figure 4.8). These are single-valued and concatenated to form 17 features.

4.2.3 Grayscale Histogram Features

For texture features, the segmented image were converted into a grayscale image and a 256-bin histogram (Figures 4.9 and 4.10) is then constructed from the silhouette pixels, which excludes background pixels, to form a representation of the grey scale distribution. Statistical moments features similar to Spampinato et al. (2010) from the histogram were computed, which was used to form greyscale histogram features. In total eight features including mean, standard deviation, skewness, kurtosis, energy, entropy, Hu's 2nd and 3rd

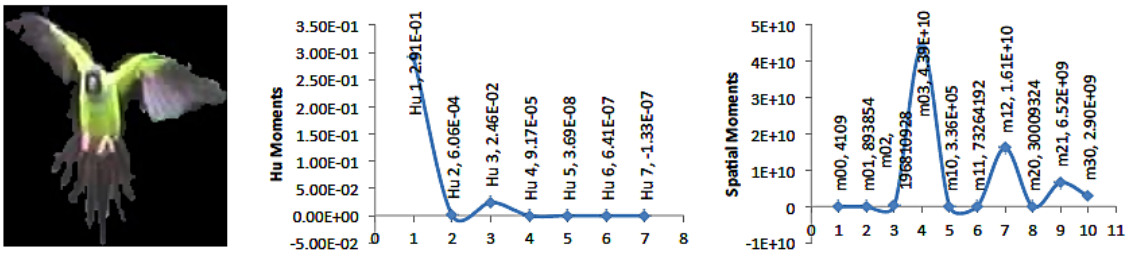


Figure 4.8: Hu (in middle) and Spatial (at right) Moments plots of a segmented Nanday Parakeet's (at left)

moments, were extracted to represent grey scale histogram features.

4.2.4 Gabor Wavelet Features

Gabor filter is used as an appearance feature in this thesis due to its salient visual properties such as spatial localisation, frequency characteristics and orientation. It also provides information about the spatial arrangement of the bird species intensities, can be used on-line and have been used by many content-based retrieval systems. According to Lee (1996), it is defined as the convolution between the function g and image $I(x,y)$, given by Equation 4.8.

$$g(x,y;\theta,\lambda,\psi,\gamma,\sigma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right) \tag{4.8}$$

Where:

- $x' = x\cos\theta + y\sin\theta$
- $y' = -x\sin\theta + y\cos\theta$
- and θ , λ, ψ, γ and σ are orientation, wavelength, phase, aspect ratio and standard deviation respectively.

Gabor wavelets features were extracted, using the following parameters $\lambda = 1$, $\psi = 0$, $\gamma = 02$, $\sigma = 1$, and $windowsize = 31$. Four different values were used for $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The result is four greyscale wavelet images (see Figure 4.11), from which a histogram was created and then five statistics were extracted: mean, standard deviation, skewness, kurtosis, energy, entropy. This yields a total of 20 features.

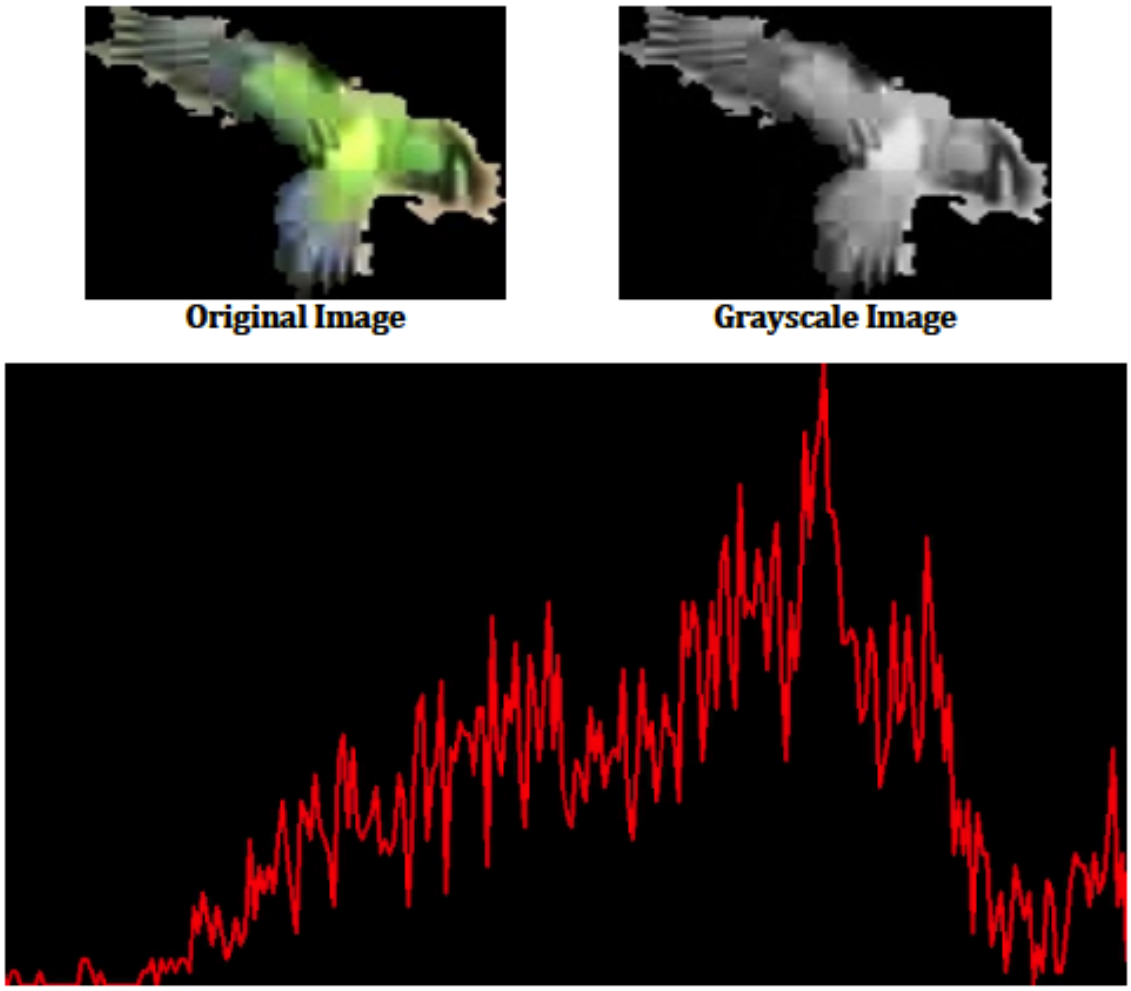


Figure 4.9: The top left image is the original image of a Cockatiel in RGB. This was transformed into grayscale (see top right of this figure) and used to form the grayscale histogram (see the bottom part of this figure). The histogram was created using the silhouette of the bird excluding background. The horizontal axis shows the intensity values from 0 to 255 and the vertical the number of intensity values.

4.2.5 Colour Log-Polar Features

Pun and Lee (2003) have demonstrated how images converted from their Cartesian plane to a log-polar image before processing, eliminate the effects of rotation and scale. Considering an image I with Cartesian coordinates denoted by $I(x,y)$. This is transformed into a log-polar form $dst(\theta,\rho)$ as in Equation 4.9. To obtain the log-polar image, a conformal mapping of the points in the Cartesian plane $I(x,y)$ to the points in the log-polar plane (θ,ρ) was performed.

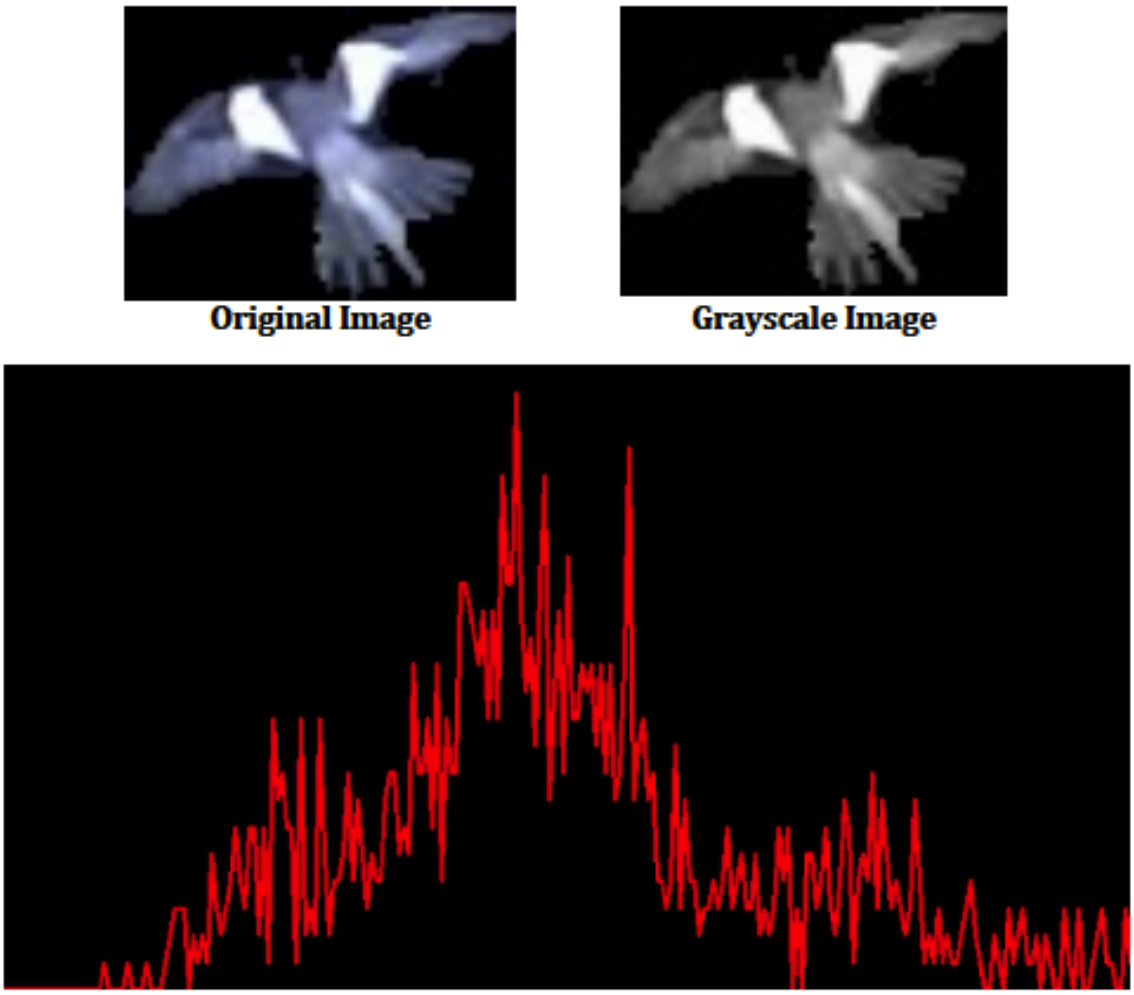


Figure 4.10: The top left image is the original image of a Cockatiel in RGB. This was transformed into grayscale (see top right of this figure) and used to form the grayscale histogram (see the bottom part of this figure). The histogram was created using the silhouette of the bird excluding background. The horizontal axis shows the intensity values from 0 to 255 and the vertical the number of intensity values.

$$dst\,(\theta,\rho) \leftarrow src\,(x,y)\,for \begin{cases} \rho = \log\sqrt{x^2+y^2} \\ \theta = \arctan\left(\frac{x}{y}\right) if x > 0 \end{cases} \tag{4.9}$$

To complement the colour features, the segmented image were converted to HSV space and a log-polar transform was applied to each channel separately (see Figure 4.12). Five statistics (including mean, standard deviation, skewness, entropy and energy) were computed from each channel. Similar to the previous process, these features were concatenated to form the log-polar feature set, which comprises a total of 15 features. This approach is different from existing approaches because it considered colour information

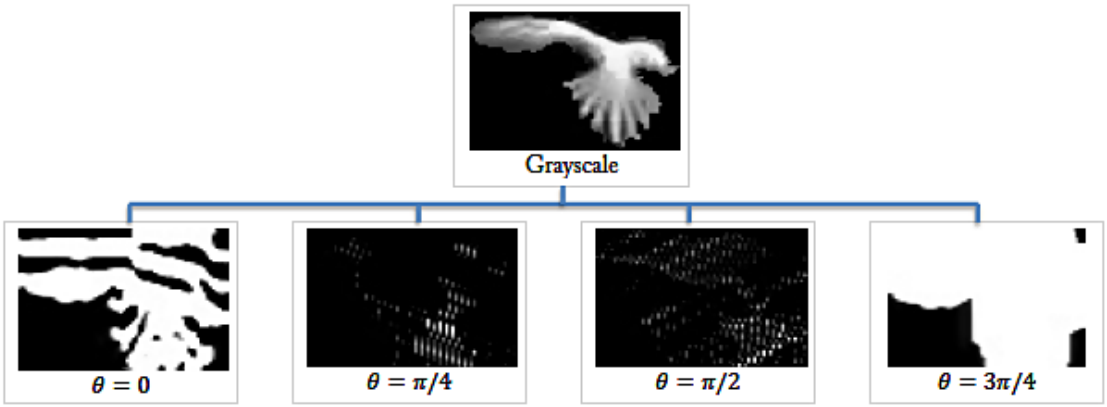


Figure 4.11: Gabor filter features for four orientations. Five Statistical features were extracted from these and the results concatenated to form a feature vector.

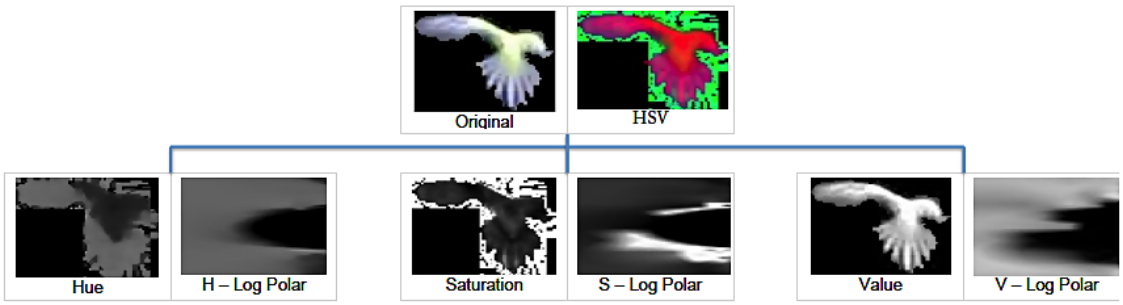


Figure 4.12: Colour log-polar features obtained by extracting hue, saturation and value log-polars then extracting statistical features, which were concatenated to form a colour log-polar feature vector.

whilst computing the log-polar features.

4.3 Experiments

Three sets of experimental evaluations were performed:

- Firstly, quantifying the effectiveness of the appearance feature set, across the dataset of seven classes (seven different species), using four different classifiers and comparing this result with the feature sets proposed by Marini et al. (2013).

- Secondly, quantifying the effectiveness of the appearance feature set, across the extended dataset of thirteen classes (eleven bird species, one with three colour forms), using four different classifiers. This was compared with the feature sets proposed by Marini et al. (2013).

- Finally, quantifying the effectiveness of the appearance feature set, across the CUB-200-2011 dataset (using 5, 17 and 200 species), using four different classifiers. Again, this was compared with the feature sets proposed by Marini et al. (2013). Features for these experiments were selected randomly.

All experiments were performed on a Mac book pro laptop running OS X 19.5, with 2.5 GHz Processor and 4 GB RAM. The pre-processing and feature extraction algorithms were all implemented in C++ with XCode 5.1.1 and OpenCV 3.0, whilst the classification and feature selection algorithms were implemented in WEKA 3.7 (Hall et al., 2009).

For all experiments, a five-fold cross-validation scheme was used, in which the original videos for each class (species) were randomly partitioned into five equal sized subsamples each. Of the five subsamples for each class (species), a single subsample was retained as the validation data for testing the model, and the remaining four subsamples were used as training data. The cross-validation process was then repeated five times (the folds), with each of the five subsamples used exactly once as the validation data (see figure 4.13 for a diagram of a thirteen class five-fold cross validation). The five results from the folds were then averaged to produce the results (correct classification rates) for each experiment. The advantage of doing it this way was that all videos were used for both training and validation, and each video was used for validation exactly once.

For each experimental run, individual image frames were sampled (from the training and test set), from which the corresponding appearance features were extracted (see Section 4.2). These features were concatenated to form the full appearance feature set, comprising 169 features. All features were stored in a WEKA compatible format.

Finally, the feature set was loaded into WEKA for classification. Four classifiers were then used (Naive Bayes (NB), Random Forest (RF), Random Tree (RT) and Support Vector Machine (SVM)) to perform the classification experiments, and results for each were reported in Section 4.4. The Naive Bayes classifier is based on the Bayes rule assuming conditional independence between classes. This is a multi-class classifier as when given an observation, the classifier estimates the conditional probabilities of classes using the joint probabilities of samples and classes and the observation is classified as

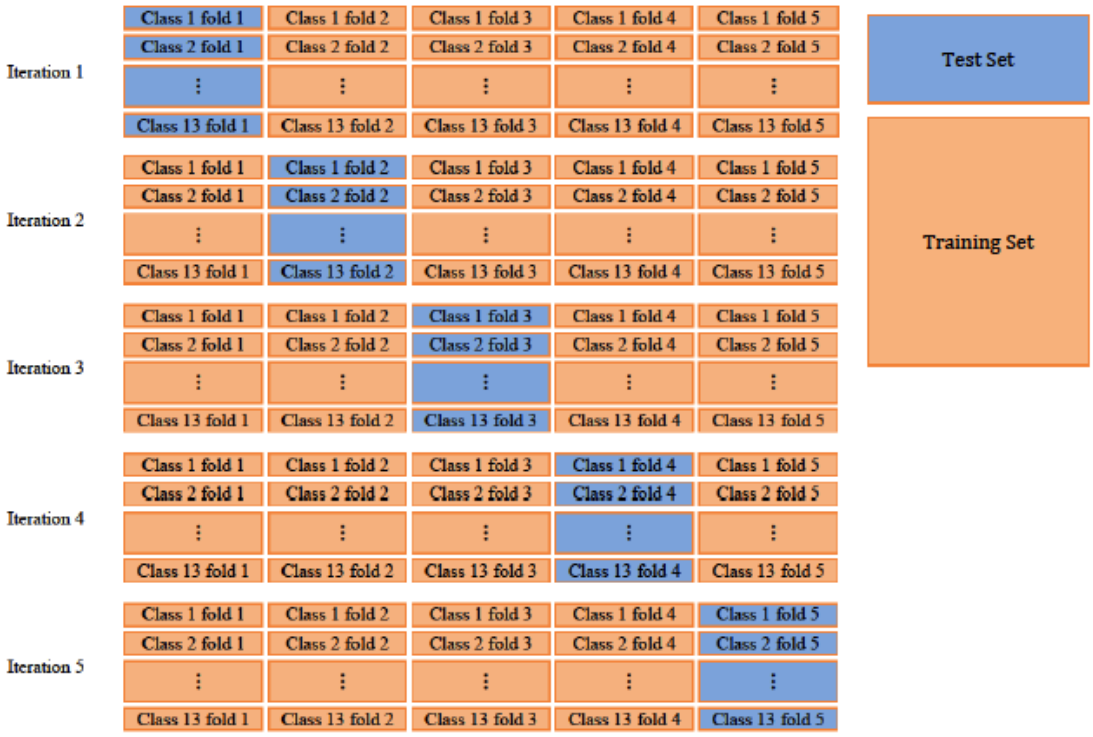


Figure 4.13: A sample five-fold cross-validation assuming the thirteen class dataset was used. Each fold of a class represents approximately one-fifths of the videos in that class. In each iteration, one fold of each of the classes is used for testing and the remaining for training.

one of the N classes based on the conditional probabilities. The SVM classifier was based on LibSVM proposed by Chang and Lin (2011), which is comparable to that used by Marini et al. (2013), and implemented using a radial basis function kernel, with the gamma and cost parameters optimised using a 5-fold grid search. In the case of the Random Tree classifier, K randomly chosen attributes at each node were considered, in this case $K = int(log_2(\#features) + 1)$, and the maximum depth of the tree was set to be *unlimited*. Twenty trees were used for the Random Forest classifier, as this results in a convergence of the *out of bag* errors (other parameters for this classifier are the same as for the RT).

The three broad methods that can be applied to compare the statistical significance of the results of two classifiers over k-folds cross-validation are the paired t-test, Wilcoxon's signed-rank test and the sign test. These methods can be applied to the results of either the k-fold cross-validation directly or r-times k-fold cross-validation. We compare the statistical significance of the results of two classifiers using the Wilcoxon's signed-rank

test applied on the r-times k-fold cross-validation results, as in Fay (2007). The results of the $r.k$ experiments performed with both learning algorithms are used as input to this statistical significance tests. In all experiments involving statistical significance of classifiers, $r = 10$ and $k = 5$ were used. The Wilcoxon's signed-rank test was chosen as it is a non-parametric alternative to the paired t-test, more powerful than the signed test (Fay, 2007) and is not substantially affected by the presence of outliers Demšar (2006).

For all statistical significance test performed, the null hypothesis H_0 is that there is no significant difference between the two classifiers being compared while the alternative hypothesis H_1 is that, there is a significant difference. All tests are carried out as a two-sided test with a 49 degree of freedom and $\alpha = 0.05$ significance level. This corresponds to a critical value of 415. The decision rule is that H_0 is rejected if the test statistic W is less than 415. W is defined as the smaller of W^+ and W^-, where W^+ is the sums of the positive ranks and W^-, the negative ranks.

4.4 Results

This section presents results of the experiments using the appearance features with the four classifiers: Naive Bayes (NB), Random Forest (RF), Random Tree (RT) and Support Vector Machine (SVM). In particular, an evaluation of the performance of the appearance features against that of Marini et al. (2013), on the three datasets (dataset #1, #2 and Caltech-UCSD Birds-200-2011) were presented.

4.4.1 Initial Results Based on the Seven Species Dataset

Tables 4.4 and 4.5 show the results (classification rates) of the appearance feature sets and that of Marini et al. (2013) respectively, using all four classifiers (NB, RF, RT and SVM).

Comparing the results of the appearance features with those used by Marini et al. (2013), it revealed that they outperformed Marini et al. (2013)'s considerably on all four classifiers. The appearance features produced consistently better results than that of Marini et al. (2013)'s colour features, including when they are used with the RF and RT classifiers. On this dataset, the appearance features have achieved between 6-13%

Table 4.4: Correct classification rates by species, based on **appearance feature sets** without feature selection using the four standard classifiers. These results use the **seven species dataset** and shows ± standard deviation across the five folds.

Species	NB	RF	RT	SVM
a=Common House Martin	93% ± 0.10%	98% ± 0.71%	88% ± 0.09%	94% ± 0.57%
b=Common Wood Pigeon	86% ± 0.54%	95% ± 0.55%	87% ± 0.48%	86% ± 0.36%
c=Superb Starling	83% ± 0.41%	92% ± 0.08%	83% ± 0.43%	86% ± 0.87%
d=Nanday Parakeet	90% ± 0.13%	96% ± 0.43%	85% ± 0.57%	85% ± 0.58%
e=Cockatiel	83% ± 0.02%	90% ± 0.13%	81% ± 0.16%	83% ± 0.28%
f=Common Starling	88% ± 0.80%	97% ± 0.74%	89% ± 0.53%	95% ± 0.08%
g=Budgerigar (wild-type)	83% ± 0.63%	89% ± 0.71%	77% ± 0.32%	70% ± 0.65%
Overall Correctly Classified	**87% ± 0.47%**	**94% ± 0.56%**	**85% ± 0.39%**	**86% ± 0.42%**

Table 4.5: Correct classification rates based on **Marini et al. (2013)'s features** without feature selection using the four standard classifiers. These results use the **seven species dataset** and shows ± standard deviation across the five folds.

Species	NB	RF	RT	SVM
a=Common House Martin	86% ± 0.10%	89% ± 0.42%	74% ± 0.03%	77% ± 0.25%
b=Common Wood Pigeon	67% ± 0.16%	87% ± 0.24%	78% ± 0.86%	82% ± 0.20%
c=Superb Starling	84% ± 0.15%	86% ± 0.27%	84% ± 0.70%	85% ± 0.32%
d=Nanday Parakeet	60% ± 0.45%	90% ± 0.38%	77% ± 0.58%	81% ± 0.06%
e=Cockatiel	70% ± 0.20%	89% ± 0.89%	81% ± 0.75%	82% ± 0.60%
f=Common Starling	89% ± 0.23%	91% ± 0.47%	76% ± 0.24%	77% ± 0.25%
g=Budgerigar (wild-type)	68% ± 0.69%	80% ± 0.09%	68% ± 0.47%	57% ± 0.59%
Overall Correctly Classified	**74% ± 0.31%**	**88% ± 0.43%**	**77% ± 0.67%**	**78% ± 0.35%**

more correct classification rates compared with Marini et al. (2013). The statistical significance of the correct classification rate of the appearance feature set was compared with that of Marini et al. by computing the Wilcoxon's test statistic $W = 0$ (the smaller of $W^+ = 571$ and $W^- = 0$). $W = W^- = 0$ since the signs of the difference between the two classifiers over the 10 times 5-fold cross validation trials are all positive. In other words, the classification rate of the first classifier is better than the second over all 50 trials. Since the computed test statistic is less than 415, we accept the alternative hypothesis that the correct classification rate of the appearance features set is significantly different from that of Marini et al. (2013)'s colour features. Hence, the assertion is accepted that the appearance feature set produces better classification rate and which is significantly different from Marini et al.'s.

Even though this outperformed Marini et al. (2013), similar trends in classification accuracies were observed. It is evident from Tables 4.4 and 4.5 that the performance of Random Forest is superior to the other three classifiers (Naive Bayes, Random Tree

and SVM). The Random Forest (RF) classifier gives the highest correct classification rate with all combinations of features. The Random Forest classifier with the appearance features showed up to 9% increase in classification rate when compared with the other three classifiers (see table 4.4) and up to 14% increase with Marini et al. (2013)'s features. In particular, the accuracies of RF classifier are 94% and 88% for the appearance features and that of Marini et al. (2013)'s respectively.

To confirm that the Random Forest (RF) classifier's correct classification rate is statistically significant compared with the other classifiers, the Wilcoxon's test statistic W is calculated. For RF compared with NB classifier, $W = 0$ (the smaller of $W^+ = 659$ and $W^- = 0$), when RF and RT are compared, $W = 0$ (this is the smaller of $W^+ = 754$ and $W^- = 0$) and $W = 0$ (the smaller of $W^+ = 694$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF classifier resulted in a W less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB classifier. This is also true for the RF compared with the RT and SVM classifiers, as both resulted in a W which is less than 415. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the seven species dataset is used with the appearance feature sets.

Tables 4.6 and 4.7 show the confusion matrices of results obtained using the appearance features and those of Marini et al. (2013) respectively, all based on the RF classifier. The maximum classification rate using the appearance features is 98% and the minimum is 89% whiles that of Marini et al. (2013) is 91% and 80% respectively. Based on the proposed method, the best classification is obtained for Common House Martain (98%) while the lowest is for Budgerigar (wild-type) (89%). Using Marini et al. (2013), the best classification rate is obtained for Common Starling (95%) and the lowest for Budgerigar (wild-type), at 80%.

Also using the features from Marini et al. (2013), segmentation error can cause classification errors. In table 4.7, Nanday Parakeets (green colour) and some bird species

Table 4.6: The confusion matrix based on the **Random Forest** classifier without feature selection and using **the appearance feature sets** with the **seven species dataset**. %CC is the percentage corectly classified.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**%CC**	**Samples**
a=Common House Martin	88.1%	1.9%	0.7%	1.5%	6.1%	0.6%	1.2%	88%	4378
b=Common Wood Pigeon	2.3%	86.8%	1.0%	4.8%	1.8%	1.3%	1.9%	87%	4347
c=Superb Starling	1.4%	2.1%	83.4%	2.0%	9.8%	0.2%	1.1%	83%	1920
d=Nanday Parakeet	1.9%	5.2%	1.0%	85.1%	1.9%	1.9%	3.1%	85%	4155
e=Cockatiel	7.1%	2.5%	4.3%	1.9%	80.8%	1.4%	2.1%	81%	3942
f=Common Starling	0.9%	1.7%	0.6%	3.1%	1.4%	89.4%	3.0%	89%	2914
g=Budgerigar (wild-type)	2.3%	3.6%	1.8%	7.2%	3.8%	3.8%	77.5%	77%	2276
Overall Correctly Classified								**85%**	

which were filmed with green backgrounds (green grass and trees) had contaminated segmentation, especially Wood Pigeons (see Fig. 4.14) and were therefore misclassified. The proposed method remains robust to these specific errors. Specifically, 2.9% more instances of Nanday Parakeets were misclassified as Common Wood Pigeons when Marini et al.'s colour features were used compared to the proposed appearance feature set.

Appearance-Related Species

The performance of the proposed method in differentiating between species of similar colouration have been examined. In this experiment, there were only two species with similar colouration, Nanday Parakeet and Budgerigar (wild-type). They have very similar colour features (Fig. 4.1), and it appears that Budgerigar (wild-type) are typically misclassified as Nandy Parakeets when relying on colour alone. When Marini et al. (2013)'s appearance features, which uses colour only, are used to classify these species, more misclassifications are recorded than when the proposed appearance features are used. For example, the misclassification of Budgerigar (wild-type) as Nanday parakeets reduced by 9.1%. This suggests that merging colour and shape features, as in the proposed appearance features, may increase the correct classification rates for species with a similar colour.

4.4.2 Results Based on the Thirteen Classes Dataset

The previous experiment in section 4.4.1 shows that the set of appearance features outperformed those of Marini et al. (2013), on the smaller dataset of seven species, using four classifiers. This section presents an expanded set of comparative results which also

Table 4.7: The confusion matrix based on the **Random Forest** classifier without feature selection and using **Marini et al. (2013)'s feature sets** with the **thirteen classes dataset (eleven bird species, one with three colour forms)**. %CC is the percentage corectly classified.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**%CC**	**Samples**
a=Common House Martin	89.3%	4.2%	0.1%	0.2%	6.0%	0.1%	0.1%	89%	4378
b=Common Wood Pigeon	6.2%	86.5%	0.0%	6.2%	0.4%	0.3%	0.3%	87%	4347
c=Superb Starling	0.4%	3.2%	85.8%	0.2%	10.2%	0.1%	0.1%	86%	1920
d=Nanday Parakeet	2.2%	2.0%	0.3%	90.2%	0.2%	0.4%	4.7%	90%	4155
e=Cockatiel	5.3%	0.9%	2.5%	0.4%	89.3%	1.1%	0.5%	89%	3942
f=Common Starling	0.0%	0.3%	0.0%	4.5%	0.3%	90.8%	4.1%	91%	2914
g=Budgerigar (wild-type)	1.9%	1.6%	0.0%	13.2%	1.4%	2.1%	79.9%	80%	2276
Overall Correctly Classified								**88%**	

Figure 4.14: Segmented Wood Pigeons contaminated with green background.

incorporate the expanded video dataset of thirteen classes (eleven bird species, one with three colour forms). Tables 4.8 and 4.9 shows the results for the expanded set of thirteen classes (eleven bird species, one with three colour forms) using our appearance feature sets and that of Marini et al. (2013)'s respectively using the four standard classifiers.

Again, it is evident from Tables 4.8 and 4.9 that the performance of Random Forest is superior to the three classifiers(Naive Bayes, Random Tree and SVM). The Random Forest (RF) classifier gives the highest correct classification rate with all combinations of features and datasets, which assert that this is the most effective classifier of the four tested for this problem domain. This is particularly evident with the larger set of thirteen classes, where the RF classifier outperforms NB, RT and SVM by 31%, 17% and 15% respectively based on the overall correct classification rate, using the appearance features. Similarly, using the Marini et al. feature set the RF classifier outperformed the NB, RT and SVM classifiers by 29%, 16% and 21% respectively.

Again, the statistical significance of the Random Forest (RF) classifier's correct classification rate compared with the other classifiers were undertaken by computing the

Table 4.8: Correct classification rates based on **the appearance features** without feature selection using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in this chapter and shows ± standard deviation across the five folds.

Species	**NB**	**RF**	**RT**	**SVM**
a=Alexandrine Parakeet	27% ± 0.63%	82% ± 0.33%	61% ± 0.44%	56% ± 0.40%
b=Nanday Parakeet	47% ± 0.83%	83% ± 0.04%	56% ± 0.49%	47% ± 0.28%
c=Blue-crowned Parakeet	48% ± 0.62%	84% ± 0.11%	68% ± 0.14%	61% ± 0.28%
d=Common House Martin	77% ± 0.46%	96% ± 0.79%	86% ± 0.69%	96% ± 0.02%
e=Eastern Rosella	53% ± 0.14%	80% ± 0.23%	54% ± 0.61%	48% ± 0.08%
f=Budgerigar (yellow)	60% ± 0.85%	85% ± 0.17%	59% ± 0.10%	71% ± 0.55%
g=House Sparrow	45% ± 0.31%	70% ± 0.31%	52% ± 0.70%	39% ± 0.30%
h=Budgerigar (wild-type)	22% ± 0.12%	60% ± 0.76%	40% ± 0.85%	23% ± 0.11%
i=Common Wood Pigeon	47% ± 0.60%	72% ± 0.85%	56% ± 0.61%	53% ± 0.18%
j=Black-headed Gull	64% ± 0.82%	98% ± 0.61%	86% ± 0.65%	95% ± 0.27%
k=Cockatiel	56% ± 0.50%	84% ± 0.21%	61% ± 0.82%	63% ± 0.28%
l=Budgerigar (blue)	35% ± 0.35%	78% ± 0.73%	56% ± 0.89%	48% ± 0.82%
m=Common Starling	52% ± 0.52%	73% ± 0.67%	53% ± 0.80%	74% ± 0.47%
Overall Correctly Classified	**54% ± 0.51%**	**86% ± 0.45%**	**68% ± 0.62%**	**70% ± 0.30%**

Wilcoxon's test statistic W. When the RF classifier was compared with NB, $W = 0$ (the smaller of $W^+ = 693$ and $W^- = 0$), $W = 0$ (this is the smaller of $W^+ = 697$ and $W^- = 0$) when RF and RT were compared and $W = 0$ (the smaller of $W^+ = 662.5$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF resulted in a $W = 0$, which is less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB. This is also true for the RF compared with the RT and SVM classifiers. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the thirteen classes dataset is used with the appearance feature sets.

Comparing these results for the proposed appearance features with those used by Marini et al. (2013) also reinforces the previous results in Section 4.4.1. The appearance features produce consistently better results than that of Marini et al. (2013) on this dataset, including when they are used with the RF classifier. On this extended dataset, the proposed appearance features have achieved between 6-14% higher classification rates compared with Marini et al. (2013), across the four standard classifiers.

Marini et al. (2013) have previously shown that increasing the number of species

Table 4.9: Correct classification rates by species, based on **Marini et al.'s features** without feature selection using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in this chapter and shows ± standard deviation across the five folds.

Species	NB	RF	RT	SVM
a=Alexandrine Parakeet	21% ± 0.74%	78% ± 0.70%	58% ± 0.84%	47% ± 0.34%
b=Nanday Parakeet	41% ± 0.87%	76% ± 0.37%	53% ± 0.75%	26% ± 0.26%
c=Blue-crowned Parakeet	39% ± 0.48%	83% ± 0.18%	64% ± 0.50%	53% ± 0.56%
d=Common House Martin	77% ± 0.28%	96% ± 0.80%	82% ± 0.25%	95% ± 0.41%
e=Eastern Rosella	47% ± 0.52%	71% ± 0.06%	49% ± 0.66%	26% ± 0.07%
f=Budgerigar (yellow)	53% ± 0.77%	73% ± 0.81%	50% ± 0.48%	49% ± 0.86%
g=House Sparrow	34% ± 0.82%	56% ± 0.18%	40% ± 0.11%	31% ± 0.64%
h=Budgerigar (wild-type)	14% ± 0.13%	43% ± 0.11%	33% ± 0.60%	16% ± 0.74%
i=Common Wood Pigeon	41% ± 0.14%	62% ± 0.55%	49% ± 0.36%	10% ± 0.65%
j=Black-headed Gull	54% ± 0.02%	93% ± 0.27%	77% ± 0.10%	87% ± 0.86%
k=Cockatiel	56% ± 0.08%	73% ± 0.08%	53% ± 0.40%	39% ± 0.77%
l=Budgerigar (blue)	32% ± 0.47%	68% ± 0.87%	51% ± 0.23%	23% ± 0.42%
m=Common Starling	44% ± 0.79%	52% ± 0.05%	40% ± 0.36%	39% ± 0.49%
Overall Correctly Classified	**48% ± 0.47%**	**77% ± 0.32%**	**61% ± 0.45%**	**56% ± 0.54%**

(unsurprisingly) reduces the correct classification rate. This is also evident in the evaluations, with a noticeable decrease in the correct classifications using the extended dataset (dataset #2), across all features sets and classifiers. However, it is also evident that the RF classifier is the most robust of the three. Across all feature sets, correct classification rates using this classifier dropped by only 9% using our feature sets, and 11% using Marini et al. (2013)'s features. The other classifiers dropped their correct classification rates by between 16-33%. Therefore, the conclusion is that the RF classifier is generally superior with the datasets and feature set tested.

Appearance-Related Species

Tables 4.10 and 4.11 show the confusion matrices of results obtained using the proposed appearance features and those of Marini et al. (2013) respectively, based on the RF classifier. The cross-species confusion matrix for the other classifiers have been presented in Appendix A. The maximum classification rate using the proposed appearance features is 98% and the minimum is 60% whilst that of Marini et al. (2013) is 96% and 43% respectively. Based on the proposed method, the best classification was obtained for Black-headed gull (98%) while the lowest is for Budgerigar (wild-type) (60%). Using Marini et al. (2013), based on the averages, the best classification rate is obtained for Common

Table 4.10: The confusion matrix based on the **Random Forest** classifier without feature selection, using **the appearance features** on the **thirteen classes dataset (eleven bird species, one with three colour forms)**. %CC is the percentage corectly classified.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	81.9%	6.9%	1.5%	0.3%	0.3%	1.2%	0.6%	1.1%	0.5%	1.6%	1.1%	2.6%	0.4%	82%	12801
b=Nanday Parakeet	7.2%	82.8%	2.2%	0.9%	0.5%	1.5%	0.2%	0.9%	1.2%	0.5%	0.8%	0.8%	0.6%	83%	10025
c=Blue-crowned Parakeet	8.1%	2.1%	84.2%	0.3%	0.9%	0.8%	0.7%	1.1%	0.2%	0.3%	0.2%	1.1%	0.2%	84%	9076
d=Common House Martin	0.1%	0.2%	0.1%	96.4%	0.0%	0.0%	1.4%	0.0%	0.0%	0.4%	0.1%	0.0%	1.2%	96%	25517
e=Eastern Rosella	1.9%	1.4%	1.9%	0.3%	80.4%	0.8%	1.2%	0.8%	0.2%	4.0%	1.6%	5.1%	0.4%	80%	5929
f=Budgerigar (yellow)	3.8%	2.7%	0.9%	0.4%	0.3%	85.4%	1.3%	1.6%	0.1%	1.6%	0.5%	1.0%	0.2%	85%	7667
g=House Sparrow	0.4%	0.1%	0.2%	11.5%	0.2%	0.8%	70.4%	0.3%	0.0%	4.2%	1.4%	1.7%	9.0%	70%	10191
h=Budgerigar (wild-type)	7.7%	4.4%	3.7%	2.5%	1.4%	6.0%	3.2%	60.4%	0.6%	3.2%	2.1%	3.4%	1.5%	60%	6283
i=Common Wood Pigeon	2.3%	7.0%	2.8%	3.6%	0.5%	0.8%	0.8%	0.6%	72.2%	1.7%	2.4%	1.0%	4.3%	72%	4301
j=Black-headed Gull	0.2%	0.1%	0.0%	0.3%	0.1%	0.0%	0.5%	0.0%	0.0%	98.2%	0.1%	0.5%	0.0%	98%	38764
k=Cockatiel	0.5%	0.3%	0.1%	4.6%	0.2%	0.2%	1.2%	0.2%	0.1%	3.2%	83.8%	1.1%	4.5%	84%	9398
l=Budgerigar (blue)	2.4%	0.5%	0.4%	1.1%	1.1%	0.7%	1.2%	1.0%	0.0%	12.1%	1.2%	77.6%	0.7%	78%	12090
m=Common Starling	0.4%	0.5%	0.2%	9.3%	0.1%	0.7%	11.4%	0.1%	0.1%	1.2%	2.7%	0.4%	73.1%	73%	9865
Overall Correctly Classified														**85%**	

Table 4.11: The confusion matrix based on the **Random Forest** classifier without feature selection, using **Marini et al. (2013)'s features** on the **thirteen classes dataset (eleven bird species, one with three colour forms)**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	77.9%	7.3%	0.8%	0.2%	0.4%	1.9%	1.3%	1.4%	0.6%	4.0%	1.0%	2.6%	0.7%	78%	12801
b=Nanday Parakeet	10.1%	75.6%	1.2%	1.5%	1.0%	2.5%	0.4%	0.9%	2.1%	1.1%	1.6%	1.5%	0.5%	76%	10025
c=Blue-crowned Parakeet	9.9%	2.6%	83.2%	0.3%	0.6%	0.6%	0.7%	0.2%	0.3%	0.5%	0.3%	0.7%	0.2%	83%	9076
d=Common House Martin	0.1%	0.3%	0.1%	96.5%	0.0%	0.0%	0.7%	0.0%	0.1%	0.8%	0.3%	0.0%	0.9%	96%	25517
e=Eastern Rosella	3.2%	4.8%	2.4%	0.2%	71.0%	1.5%	1.1%	1.3%	0.8%	6.2%	2.0%	5.1%	0.5%	71%	5929
f=Budgerigar (yellow)	9.2%	4.5%	2.9%	0.8%	0.5%	72.7%	1.4%	2.2%	0.4%	2.4%	0.9%	1.4%	0.8%	73%	7667
g=House Sparrow	1.9%	0.3%	0.5%	13.0%	0.3%	0.6%	55.8%	0.9%	0.1%	7.3%	2.8%	1.8%	14.7%	56%	10191
h=Budgerigar (wild-type)	11.0%	5.9%	5.8%	2.5%	1.7%	7.5%	5.2%	43.2%	1.3%	7.4%	3.1%	3.7%	1.7%	43%	6283
i=Common Wood Pigeon	3.4%	10.0%	3.3%	5.2%	1.2%	1.5%	0.9%	1.1%	62.5%	3.2%	3.4%	1.0%	3.3%	62%	4301
j=Black-headed Gull	0.4%	0.2%	0.1%	1.1%	0.2%	0.1%	1.5%	0.1%	0.1%	92.6%	1.3%	1.6%	0.7%	93%	38764
k=Cockatiel	1.0%	0.8%	0.1%	7.5%	0.4%	0.6%	1.7%	0.4%	0.2%	7.9%	73.0%	1.2%	5.2%	73%	9398
l=Budgerigar (blue)	3.1%	1.1%	0.7%	1.1%	1.5%	0.8%	1.4%	1.0%	0.2%	18.8%	1.5%	68.1%	0.6%	68%	12090
m=Common Starling	1.9%	0.8%	0.3%	17.0%	0.2%	1.0%	14.7%	0.3%	0.2%	7.5%	3.6%	0.4%	52.3%	52%	9865
Overall Correctly Classified														**77%**	

House Martin (96%) and the lowest for Budgerigar (wild-type), at only 43%.

This is further motivated by an examination of the fine-grained performance of the proposed method; that is, differentiation between species of similar appearance. Given that the RF classifier has shown best overall performance, a detailed cross-species confusion matrix for this classifier was provided, obtained using the proposed appearance features with the larger dataset, shown in Table 4.10. Figure 4.15 shows examples of the Alexandrine parakeet, Nanday parakeet and Blue-crowned parakeet, which have closely related colouration.

When the proposed appearance features are used to classify these species the misclassification are reduced than when the colour features in Marini et al. (2013) are used. For example, the misclassification of Alexandrine parakeets as Nanday Parakeets reduce by 0.5% instances; Nanday Parakeets as Alexandrine Parakeets by 3.0%; and Blue-crowned Parakeets as Alexandrine Parakeets by 1.8%. This suggests that merging colour and shape

Figure 4.15: Species with closely related appearances. From left to right: Alexandrine Parakeet, Nanday Parakeet and Blue Crowned Conure

features, as in the proposed appearance features may increase the correct classification rates for species with closely-related appearances.

Motion-Related Species

The three Budgerigar species have different colour form, thus the low misclassification among these species when only appearance features were used. For example, 1.6% and 1.0% Budgerigar (yellow) were misclassified as Budgerigar(wild-type) and Budgerigar (blue) respectively, 6.0% and 3.4% Budgerigar (wild-type) were misclassified as Budgerigar (yellow) and Budgerigar (blue) respectively, and 0.7% and 1.0% Budgerigar (blue) were misclassifies as Budgerigar (yellow) and Budgerigar (wild-type) respectively.

4.4.3 Results Based on Caltech-UCSD Birds-200-2011 Dataset

The results in this section were based on the SVM classifier to enable a comparison of the results to that of Marini et al. (2013), which used the same setup. The results were presented in table 4.12. Results based on Marini et al. (2013)'s features uses various fusion rules which include Maximum (MAX), SUM (SUM), Product (PROD), Weighted Sum (WSUM), Weighted Product (WPROD). Marini et al. (2013) first classified species using each of the three colour channels of the HSV colour space. The results of each of the channels were then merged into one, to determine the class of a particular species, using the fusion rules mentioned above. Another approach that was used by Marini et al. is to merge the HSV features into one feature set for classification. The results of this using SVM classifier on 2, 5, 17 and 200 species have also been presented in table 4.12.

Again, comparing the results for the appearance features with those used by Marini

Table 4.12: The appearance features versus Marini et al. (2013)'s on the Caltech-UCSD Birds-200-2011 Dataset. Marini's HSV results are based on concatenating the hue, saturate and value features as one feature set for classification. Marini's Fusion Rules are based on the using hue, saturate and value as three different features and the results combined using the fusion rules.

#Species	Proposed Method	Marini's HSV	Marini's Fusion Rules				
			MAX	SUM	PROD	WSUM	WPROD
2	97.53%	92.64%	85.29%	86.76%	88.24%	89.71%	91.18%
5	64.22%	48.34%	47.68%	49.01%	49.67%	51.66%	51.66%
17	38.24%	25.63%	19.65%	22.16%	22.54%	23.89%	23.7%
200	11.36%	8.60%	6.76%	7.16%	7.25%	7.59%	8.03%

et al. on the Caltech-UCSD Birds-200-2011 dataset re-confirms the results in the previous sections. the proposed appearance features produce better correct classification rates than that of Marini et al. (2013). Specifically, they produced approximately 5-12% more classification rates than Marini et al.'s when 2 species were used, 12-16% with 5, 13-17% with 17 and 3-4% with 200 species (see table 4.12).

Marini et al. have shown that increasing the number of species (unsurprisingly) reduces the correct classification rate. This is also evident in this evaluation, with a noticeable decrease in the correct classifications rates across all methods used in this experiment when the number of species are increased. Across all the methods used, correct classification rates dropped when the number of classes (species) is increased. In the proposed appearance features, there was a drop in classification rate by 33% when the number of classes are increased from 2 to 5 species, 26% when increased to 17 species, 27% when increased to 200. Similarly, there was a drop in classification rate when Marini et al. (2013)'s methods are used. Specifically, there was 44% drop in classification rates when species were increased from 2 to 5, 23% when increased to 17 species, 17% when increased to 200 with Marini et al. (2013)'s best feature set (the merged HSV features). Therefore, the conclusion is that increasing the number of classes whiles using only appearance features may result in a drop in the correct classification rates.

4.5 Conclusion

The appearance feature set worked well when compared with Marini et al.'s on all four standard classifiers. All datasets and methods performed better with the RF classifier

than the other classifiers. We have shown that this classifier's correct classification rate is statistically significant compared with the other classifiers using the Wilcoxon's sign ranked test. On the seven classes dataset, the proposed appearance features achieved between 6-13% higher classification rates compared with Marini et al.'s. However, on the thirteen classes dataset, the proposed appearance features achieved between 6-14% more correct classification rates. Finally on the Caltech-UCSD Birds-200-2011 dataset, again, the proposed method outperformed Marini et al.'s by approximately between 5-12% using 2 species, 12-16% using 5, 13-17% using 17 and only 3-4% using 200.

The misclassification of bird species was due to illumination and similar colour patterns in some species. Marini et al.'s method uses only colour features and therefore had more misclassifications with similar colour patterned species than the proposed. For the seven classes dataset, the misclassification of Budgerigar (wild-type) as Nanday parakeets reduces by 9.1% when the proposed method was used. Similarly, with the extended dataset, misclassification of Alexandrine parakeets as Nanday parakeets reduced by 0.5%; Nanday parakeets as Alexandrine parakeets by 3.0%; and Blue-crowned parakeets as Alexandrine parakeets by 1.8%, when the proposed appearance method was used. Finally, due to the distinct colour features of the three Budgerigar species, misclassification among them is low.

Marini et al. (2013) have shown that increasing the number of species reduces the correct classification rate. Classification rates dropped significantly when moving from seven to thirteen classes and whilst the RF classifier remains effective, the result is a reduction of approximately 10% in the correct classification rates. When the number of classes is increased from 2 to 5, 17 and 200, the classification rates dropped with the proposed method and that of Marini et al.. Therefore, the conclusion is that increasing the number of classes whilst using only appearance features may result in a drop in the classification rates, irrespective of the appearance features used.

The next chapter will identify relevant motion features which can be extracted from video of birds in flight, and use them to classify species automatically. The appearance and motion features will be effectively combined and evaluated to determine whether this

combination gives better results than appearance or motion features alone.

Chapter 5

Classification of Bird Species using Motion Features

In the previous chapter, species were classified using appearance features and the results compared with the state-of-the-art method proposed by Marini et al. (2013) using the three datasets. Experimental results revealed that the proposed appearance features outperformed Marini et al.'s on all three datasets using all four standard classifiers (Naive Bayes (NB), Random Forest (RF), Random Tree (RT) and Support Vector Machine (SVM)). In particular, using the random forest classifier, the proposed method greatly improved correct classification rates over that of Marini et al. (2013) by about 6% on the seven species dataset and 9% on the thirteen classes dataset. There was also an improvement in correct classification when using the CUB-200-2011 dataset, by approximately 6-12% using 2 species, 12-16% using 5, 13-17% using 17 and 2 - 4% using 200 species.

The methods used in Chapter 4 use single images and appearance-based models for classification; however, bird species also exhibit distinguishing behaviours (flying, moving, poses, etc) which could also be used to help robust automated identification. This is particularly relevant to the identification of birds in flight, especially at a distance where appearance-based features such as colour tend to attenuate, whilst motion-features remain discernible. The aim of the work presented in this chapter is to investigate the potential of motion-based features for differentiation of species with closely related appearances, and also to determine whether motion and appearance features can be merged to produce

results which exceed either set alone. Firstly, a richer feature set based on motion is introduced, and use to determine whether they can classify species across the two datasets introduced in the previous chapter. In particular, motion features were investigate to determine whether they can discriminate between species with similar appearances (that is, species which were less well differentiated using appearance features in Chapter 4). Motion and appearance features were then fused and using standard classifiers, determined whether these combination is more effective than either set alone. This chapter is structured into the following sections:

- In Section 5.1 the datasets used and the processing techniques applied before motion feature extraction were introduced.
- The set of motion features used for all the experiments in this chapter are described in Section 5.2.
- In Section 5.3 the experimental work is described.
- Results from experimental work, including the motion feature and full feature set, and evaluations to determine real time performance of all the models are presented in Section 5.4.
- Finally, conclusions are drawn to the chapter in Section 5.5, which include summarising all results in the chapter and introducing briefly what will be described in the next chapter.

5.1 Datasets, Methods and Preprocessing

The extended dataset detailed in Chapter 4 have been used for all experiments presented in this chapter. As a reminder, this is "Dataset #2", which is an extended set of videos covering thirteen classes made up of eleven bird species, one (Budgerigar (*Melopsittacus Undulatus*)) with three colour forms.

For each video, appearance features are calculated per frame starting from the 64^{th} frame, while motion features are calculated using 64 frames, in strides of one frame. The

first set of motion features from a video is calculated using the first 64 frames and this is merged with the appearance feature from the 64^{th} frame, to form the first combined feature of that video. Thus, for experiments in this chapter and beyond, videos that are shorter than 64 frames are not included in the dataset. Therefore, all experiments performed using the combined or motion features, had fewer videos than those performed using appearance. Likewise, the number of images in these experiments (experiments using the combined or motion features) are also fewer, as the first 63 frames are not used in the computation of appearance features that are merged with the motion to form the combined set (see Table 5.1).

Table 5.1: Table showing the number of videos and images in thirteen classes dataset when features are combined. There are fewer videos and images when compared with the original dataset used to perform appearance features only experiments.

Species	# of videos	# of images
Alexandrine Parakeet	77	7,845
Nanday Parakeet	59	6,246
Blue-crowned Parakeet	58	5,332
Common House Martin	114	17,896
Eastern Rosella	40	3,247
Budgerigar (yellow)	47	4,329
House Sparrow	74	5,318
Budgerigar (wild-type)	41	3,349
Common Wood Pigeon	35	2,027
Black-headed Gull	142	29,695
Cockatiel	58	5,687
Budgerigar (blue)	76	7,030
Common Starling	71	5,392
Total	**892**	**103,393**

Before extracting the motion features, the centroids of the segmented bird silhouette are first extracted from each frame of the videos, after performing the pre-processing

described in Chapter 4. The 2D centroid positions are used to form a trajectory in the image frame. For any bird tracked throughout N frames, such a trajectory is described as:

$$T = \{(x_1, y_1) \dots (x_N, y_N)\} \tag{5.1}$$

j is the frame index represented as $j = \{1, \dots, N\}$, and T represents the entire trajectory of a bird, represented as a series of x and y coordinates of the centroid in the image frame. Using the entire trajectory (Eqn. 5.1) for each video, shorter overlapping sub-trajectories t_k were defined, which starts on frame k of the video, and k ranges from 1 to $N - Q + 1$; where Q is the window size:

$$t_k = \left\{(x_k, y_k) \dots (x_{k+Q-1}, y_{k+Q-1})\right\} \tag{5.2}$$

In this case $Q = 64$. The overlapping windows are in steps of a one-time frame. For example, in a video of size $N = 120$ frames, the first window starts from frame $1 \dots Q = 64$, the second window started from frame $2 \dots Q + 1 = 65$ and so on. In general terms, the k window starts from time frame $k \dots k + Q - 1$. The total number of short overlapping trajectories in this example will, therefore, be $N - Q + 1 = 57$. A box filter (Gonzalez and Woods, 2002) with a 1 x 3 kernel was then applied to reduce the effect of noise in the trajectory. The idea of using the box filter is simply to replace each trajectory value with the mean value of its neighbours, including itself. The box filter is usually thought of as a convolution filter. Like other convolutions it is based around a kernel, which represents the shape and size of the neighbourhood to be sampled when calculating the mean. Given the trajectory t_k, the smoothed trajectory st_k is the convolution of the kernel ker and the trajectory t_k given by equation 5.3

$$st_k = ker * t_k \tag{5.3}$$

Where the kernel ker is $\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$. The purpose of this filter is to attenuate the

noise (reduces the variance), leading to a more accurate estimate of the trajectory. Gaussian filter (Gonzalez and Woods, 2002) can also be applied for noise smoothing in trajectories, and is similar to the box filter, but uses a kernel that represents the shape of a Gaussian. A simple form of the Gaussian filter is a box filter. The motion features were then extracted from the set of smoothed short trajectories (see Eqn. 5.2) to form a feature sequence which were used for classification. The motion features extracted are described in the following section.

5.2 Extracting the Motion Features

This section describes in detail all the motion features used for experiments throughout this thesis. A description of how these motion features were extracted is presented, including motivation for their use and examples of where they have been used. They include curvature scale space (CSS), turn based, wing-beat frequency, centroid distance function, vicinity and curvature based on sine and cosine.

5.2.1 Curvature Scale Space

Curvature scale space (CSS) is rotation and translation invariant and has been shown to be effective in distinguishing object trajectories by their concave and convex shapes (Mokhtarian et al., 1996; Beyan and Fisher, 2013b; Mai et al., 2010; Bashir et al., 2006). They have also been shown to be robust in the presence of noise. Mai et al. (2010) have shown that CSS can be effective for matching and recognizing shapes which are distorted by affine transforms, including translation, rotation, and scaling.

This was considered an important feature for two reasons. Firstly, the datasets consist of birds flying in different directions and orientations, therefore similar trajectories may appear as different relative orientation with respect to the optical axis. Secondly, they may appear at different scales due to varying distance from the camera.

Consider the parametric vector equation for the trajectory: t(u)=(x(u), y(u)), where u is some parameter. Then the curvature $\kappa(u)$ at every point on the trajectory is calculated using equation 5.4, where $x^{'}(u)$, $x^{''}(u)$, $y^{'}(u)$ and $y^{''}(u)$ are first and second derivatives of $x(u)$ and $y(u)$ respectively:

$$\kappa(u) = \frac{x^{'}(u)y^{''}(u) - x^{''}(u)y^{'}(u)}{\left(x^{'}(u)^2 + y^{'}(u)^2\right)^{\frac{3}{2}}} \tag{5.4}$$

To compute these features, each component of the sub-trajectories t_k are first convolved with a 1D Gaussian smoothing kernel $g(u,\sigma)$ (see equation 5.8) of width σ. The resulting smoothed curve becomes $t(\sigma)$ with components $X(u,\sigma)$ and $Y(u,\sigma)$ given by:

$$\begin{aligned} X(u,\sigma) &= x(u) * g(u,\sigma) \\ Y(u,\sigma) &= y(u) * g(u,\sigma) \end{aligned} \tag{5.5}$$

Now using the properties of convolution, the first and second derivatives can be calculated as in equation 5.6:

$$X_u(u,\sigma) = x(u) * g_u(u,\sigma) \tag{5.6a}$$

$$X_{uu}(u,\sigma) = x(u) * g_{uu}(u,\sigma) \tag{5.6b}$$

$$Y_u(u,\sigma) = y(u) * g_u(u,\sigma) \tag{5.6c}$$

$$Y_{uu}(u,\sigma) = y(u) * g_{uu}(u,\sigma) \tag{5.6d}$$

Since the first and second derivatives of the shorter trajectories are known from equations 5.6a to 5.6d, the curvature $\kappa(u,\sigma)$ on $t(\sigma)$ can be computed.

$$\kappa(u,\sigma) = \frac{X_u(u,\sigma)Y_{uu}(u,\sigma) - X_{uu}(u,\sigma)Y_u(u,\sigma)}{(X_u(u,\sigma)^2 + Y_u(u,\sigma)^2)^{\frac{3}{2}}} \tag{5.7}$$

The CSS motion features were then computed by iteratively applying a Gaussian smoothing kernel (see equation 5.8) with different standard deviations ($\sigma_i = 1.0 + 0.1(i)$

for $i = 0 \ldots 120$), to each sub-trajectory t_k. At each level of standard deviation, a corresponding zero crossing from the second derivative of the trajectory is recorded to form the CSS image as in Mokhtarian et al. (1996). Figure 5.1 shows examples of CSS images extracted from the trajectories of Alexandrine Parakeet, Nanday Parakeet, Common Starling and Budgerigar (wild-type). These consist of arch-shaped contours which represents the inflection points of the shape of the trajectory as it is smoothed. This process continues until there are no zero crossings and the trajectory becomes a convex curve (the chosen value for i allows the trajectories to become a convex curve).

$$g(u,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{-u^2}{2\sigma^2}} \tag{5.8}$$

The CSS feature set (represented in Equation 5.10) is formed by extracting the ten statistics in Equation 5.9 from the absolute curvature (κ) in Equation 5.4. These are concatenated with the number of curves (*curves*) in the CSS image, the total length of all the curves (*lenC*) in the CSS image and ten statistics computed from the CSS maxima (peaks) using Equation 5.9. The details of the ten statistical features have been presented in Chapter 4.

$$\mathcal{S} = \{\mu, \sigma, skew, kurt, ent, min, max, lmin, lmax, \zeta\} \tag{5.9}$$

Where μ is mean, σ is standard deviation, *skew* is skewness, *kurt* is kurtosis, *ent* is entropy, *min* is minimum, *max* is maximum, *lmin* is local minima, *lmax* is local maxima and ζ is zero crossings. All of these values were computed over κ (see Equation 5.10).

$$css_{k,j} = \{curves_{k,j}, lenC_{k,j}, \kappa^{\mathcal{S}}_{k,j}, cssMaxima^{\mathcal{S}}_{k,j}\} \tag{5.10}$$

Where $\kappa^{\mathcal{S}}$ is the ten statistics derived from κ, $cssMaxima^{\mathcal{S}}$ is the ten statistics derived from the css maximas and *css* is the sequence of Curvature Scale Space features used as part of the motion feature set. The subscripts j and k represents the videos in the dataset and the frames of these videos respectively.

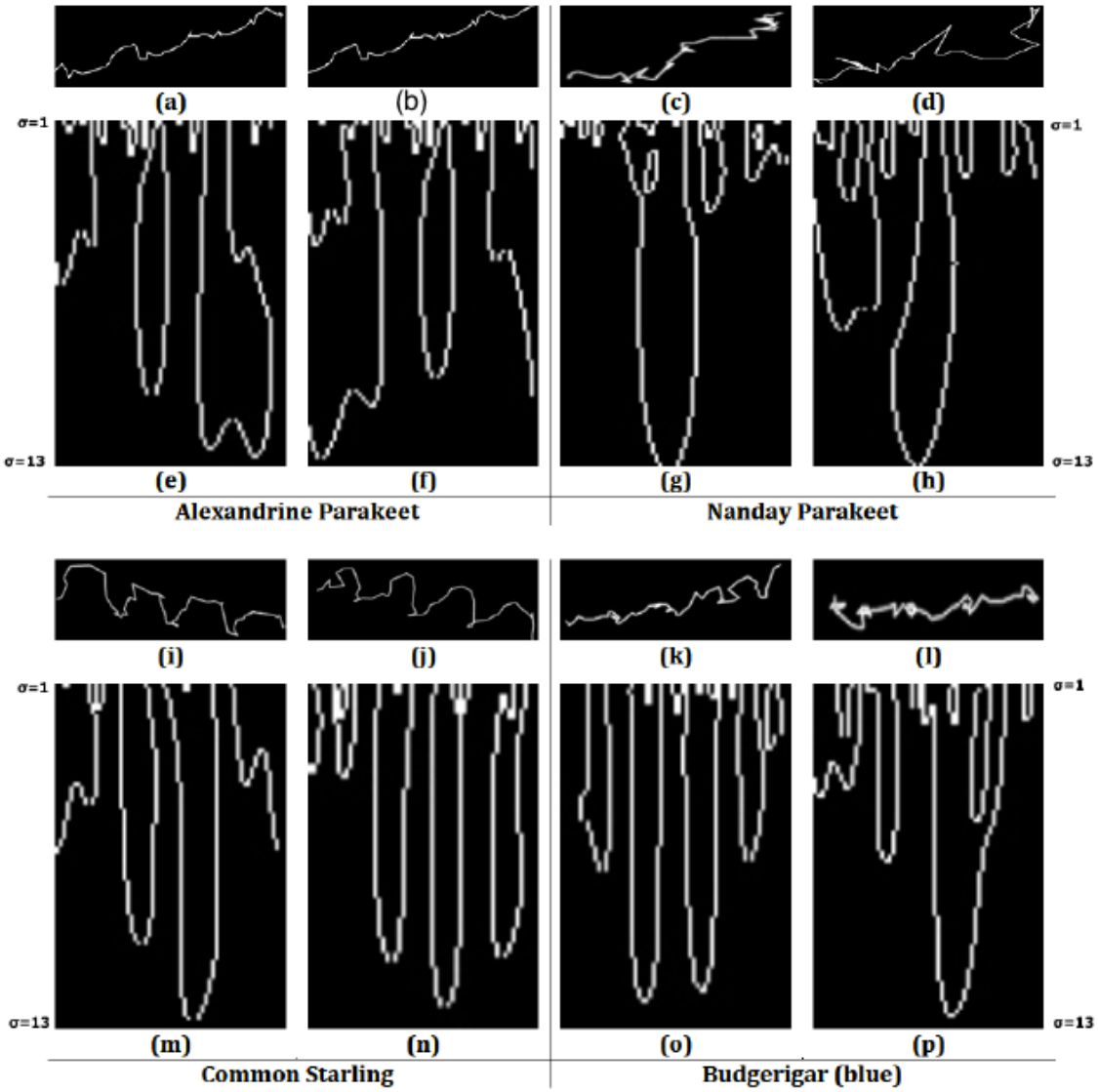


Figure 5.1: Figure (a) and (b) are sample flight trajectories of Alexandrine Parakeet with (e) and (f) being their corresponding CSS images. Figure (c) and (d) are sample flight trajectories of Nanday Parakeet with (g) and (h) being their corresponding CSS images. Figure (i) and (j) are sample flight trajectories of Common starling with (m) and (n) being their corresponding CSS images. Figure (k) and (l) are sample flight trajectories of Budgerigar (wild-type) with (o) and (p) being their corresponding CSS images. The vertical axis of all CSS plots ranges from $\sigma = 1$ (top) to $\sigma = 13$ (bottom)

5.2.2 Turn Based Features

Li et al. (2006) used turn-based features to form trajectory directional histogram, which was used to describe the statistic directional distribution of vehicle trajectories. The features were then used for categorisation, achieving an overall accuracy of 99%. Recently, Beyan and Fisher (2013a) have used turn-based features as part of a large feature set to successfully categorise abnormal fish trajectories. Turn-based features was used as part

of a feature set due to the unpredictable movement made by fish. Birds like fish also have unpredictable movements, thus it is believed this can be used as part of the feature set for classification of species.

In order to obtain the shape of each bird flight trajectory turnings were calculated. They are calculated by finding the slope of bird trajectory between two consecutive frames as given in Li et al. (2006) for vehicle trajectory classification and in Beyan and Fisher (2013b) for fish trajectory clustering. From figure 5.2, based on trajectory t, the turn θ_k between the points $P_k, P_k and P_{k+1}$ is given by Equation 5.11, used to calculate the turns in flight trajectories.

$$\theta_k = \begin{cases} arctan(\frac{y_k - y_{k-1}}{x_k - x_{k-1}}), & \text{if } (x_k - x_{k-1}) > 0. \\ arctan(\frac{y_k - y_{k-1}}{x_k - x_{k-1}}) + \pi, & \text{if } (x_k - x_{k-1}) \leq 0, (y_k - y_{k-1}) \geq 0. \\ arctan(\frac{y_k - y_{k-1}}{x_k - x_{k-1}}) - \pi, & \text{if } (x_k - x_{k-1}) \leq 0, (y_k - y_{k-1}) < 0. \end{cases} \tag{5.11}$$

Where $(x_k - x_{k-1})^2 - (y_k - y_{k-1})^2 \neq 0$, $k = 2 \ldots Q-1$ and $Q = 64$ is the length of the short trajectory defined in Section 5.1. These turn values were used to form a 62-feature set (see Equation 5.12).

$$trn_{k,j} = \{\theta_{k,j}\} \tag{5.12}$$

Where $\theta_{k,j}$ are the values computed in Equation 5.11 over $k = 2 \ldots Q-1$ and j being the video number, and $trn_{k.j}$ is the sequence of turn features used as part of the motion feature set, which is also computed over $k = 2 \ldots Q-1$ and j being the video number.

5.2.3 Wing Beat Frequency Based features

Periodic motion features associated with wing beats have been shown by Lazarevic et al. (2008) to vary among species, and this was proposed by us to provide a useful discriminating feature for classification of bird species. This includes not only short-scale features (frequency while flapping), but also others: for example some species characteristically

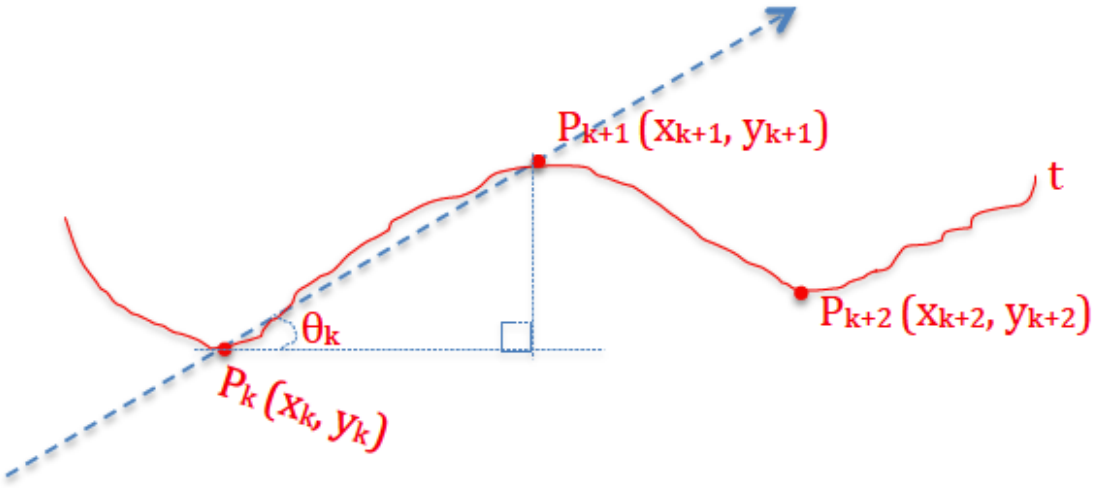


Figure 5.2: The figure shows how the slope θ_k is measured given the two points P_k and P_{k+1}.

mix flapping and gliding (Ainley et al., 2015). It has been shown in Chapter 3 that for bat species a bounding box fitted to the silhouette of a tracked individual can be used to accurately measure such periodicity features. This method has been demonstrated to outperform the state-of-the-art by Cutler and Davis (2000) when compared with the ground truth. By using these methods mean correct recognition rates of 80%, 48% and 60% were achieved when the hypothenuse, width and height metrics were used respectively. Even though it has been shown in the literature (Duberstein et al., 2012; Cullinan et al., 2015) that the wingbeat of bats and birds can sometimes help in differentiating species, this work has also confirmed this. The same method was used to extract bird periodic motion as a 1D signal broken into short overlapping windows to cover the three metrics, height, width and diagonal of the bounding box. The Fast Fourier Transform (FFT) using equation 5.13 for each of the three time signals, represented as *f(x)* were computed and the first nine frequencies from the FFT excluding DC component were extracted.

$$F(k) = \sum_{t=0}^{(N-1)} f(t)\, e^{-i2\pi kt/N} \tag{5.13}$$

Where $f(t)$ is the signal in the spatial domain with N samples, $t = 0 \ldots N-1$ and $F(k)$ in the Frequency Domain (encoding both amplitude and phase) with $k = 0 \ldots N-1$. In this case $N = 128$, which is the 64 frame short trajectory with 64 zero padding. The

effect of the zero padding is to increase the resolution of the computed frequencies, as this may help discriminate species with closely related frequencies. A larger N should be used if the samples taken are made up of larger trajectories as this may increase the resolution and help differentiate tracks much better.

The frequencies extracted from the three signals were concatenated to form a feature vector of size 27 to represent wing beat frequencies (see Equation 5.14).

$$fft_{k,j} = \{fft_{h_{k,j}}, fft_{w_{k,j}}, fft_{d_{k,j}}\} \tag{5.14}$$

fft_h represents the top nine frequencies in order of magnitude with the largest first when the height metric is used as the signal, fft_w the top nine frequencies in order of magnitude with the largest first when the width metric is used as the signal and fft_d the top nine frequencies in order of magnitude with the largest first when the hypotenuse metric is used. These are all over the k frames and j videos in the dataset. The nine frequencies were computed using Algorithm 7 and this excludes the DC component. For how the frequencies are selected see Figure 5.3, which has the top nine frequencies marked with labels.

Algorithm 7: Finding the top nine peak magnitude and corresponding frequency of the FFT

1 $MAGNITUDES \leftarrow \sqrt{\Re^2 + \Im^2}; \forall$ FFT output bin.;
2 $INDEX \leftarrow$ Find index of top nine MAGNITUDES excluding DC component.;
3 **foreach** $INDEX \leftarrow idx$ **do**
4 $\quad FREQUENCY \leftarrow idx\frac{F_S}{Q}$ Where F_S is the video frame rate (240 H_Z)
5 **end**
6 Return FREQUENCY;

5.2.4 Centroid Distance Function (CDF)

Bashir et al. (2006) have showed that using CDF (also known as shape signature) for categorisation of features results in better recognition rates compared with CSS features. Recently, this has also been used by Beyan and Fisher (2013b) as part of their feature set to discriminate abnormal from normal fish trajectories. CDF is an invariant representation

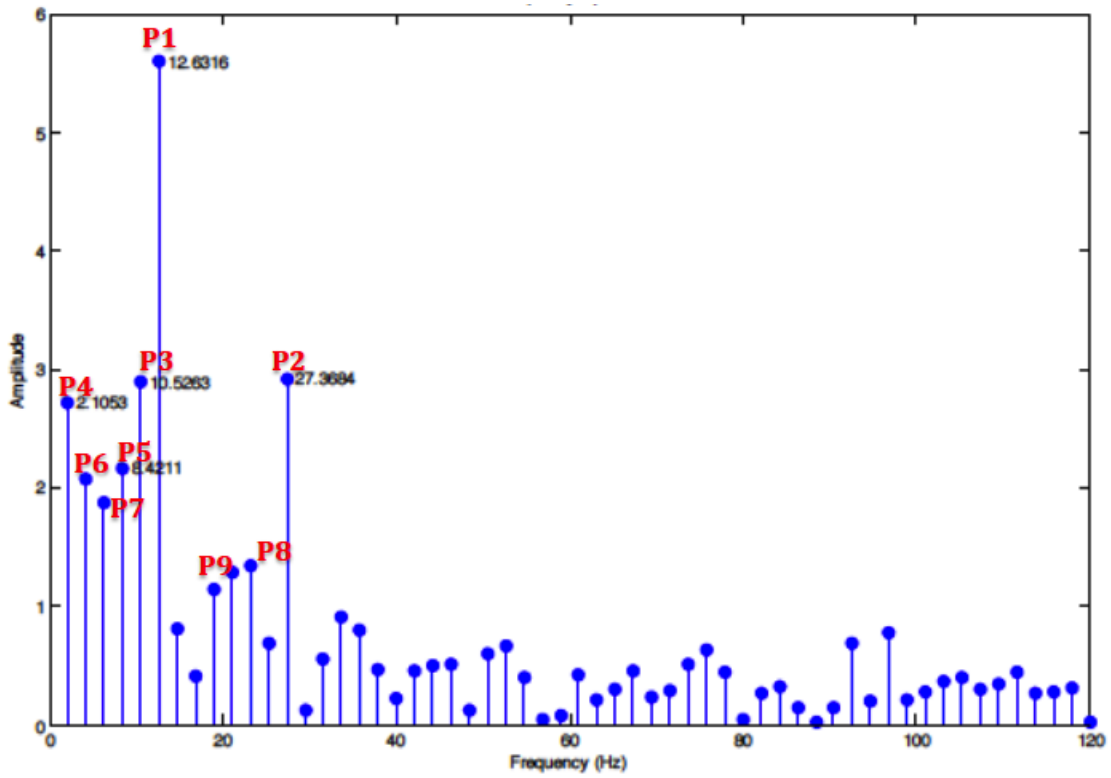


Figure 5.3: Figure the top nine frequencies in order of magnitude, with the largest first. This have been marked with red labels on the figure with $P1$ being the first selected frequency and $P9$ the last.

of the shape of data (Beyan and Fisher, 2013b; Bashir et al., 2006).

In this thesis CDF features were used to represent bird trajectories as they are not subject to rotational deformation. That is when the trajectory shape is rotated, the shape representation extracted using CDF remains the same. CDF was calculated by finding the centre point of the trajectory and calculating the distance of each trajectory point from this centre, as given by equation 5.15.

$$cdf_k = \sqrt{(x_k - x^c)^2 + (y_k - y^c)^2} \tag{5.15}$$

Where: $x^c = \frac{1}{N}\sum_{i=1}^{64} x(i)$, $y^c = \frac{1}{N}\sum_{i=1}^{64} y(i)$ and $k = 1 \ldots 64$.

To represent the CDF feature set, ten statistical features presented in Equation 5.9 were calculated and which were concatenated to form a 10-CDF feature set (see Equation 5.16).

$$cdf^{*}_{k,j} = \{\mathcal{S}^{cdf}{}_{k,j}\} \tag{5.16}$$

Where $\mathcal{S}^{cdf}$ is the ten statistics derived from cdf and cdf^{*} is the sequence of CDF features used as part of the motion feature set over all videos j and the all frames in those videos k.

5.2.5 Vicinity

Vicinity features were used by Liwicki et al. (2006) for handwriting recognition, and recently by Beyan and Fisher (2013b) for classifying abnormal fish trajectories. In both cases, the feature set was made up of vicinity aspect, curliness, slope and linearity. They were selected to represent part of the motion features, since they consist of features extracted from each point and takes into consideration their neighbouring points and are very robust to noisy data.

In this thesis a three point vicinity, which is defined as the point p_k and its preceding (p_{k-1}) and succeeding (p_{k+1}) points enclosed by a bounding box was used (see Figure 5.4). In this case $k = 1 \ldots Q$ and $Q = 64$, which is the length of the short trajectory. Four vicinity features were extracted, which include curliness, aspect, slope and linearity.

Vicinity curliness (C_k) is the total length of the trajectory in the vicinity ($L_k = |p_{k-1}p_k| + |p_k p_{k+1}|$) divided by $\max\{\delta x_k, \delta y_k\}$ as in Equation 5.17. Where $\max\{\delta x_k, \delta y_k\}$ $= \max\{(x_k - x_{k-1}), (x_{k+1} - x_k), (y_k - y_{k-1})$ and $(y_{k+1} - y_k)\}$, $|p_{k-1}p_k|$ is the distance from the point p_{k-1} to p_k and $|p_k p_{k+1}|$ the distance from p_k to p_{k+1}.

$$C_k = \frac{L_k}{\max\{\delta x_k, \delta y_k\}} \tag{5.17}$$

The ten statistical features as in Equation 5.9 were computed from the vicinity curliness features to form the curliness features set (see Equation 5.18). Where $curliness_{k,j}$ is the curliness feature set index over k frames and j videos, and $\mathcal{S}^{C_{k,j}}$ is the ten statistical features derived from the computed vicinity curliness.

$$curliness_{k,j} = \{\mathcal{S}^{C_{k,j}}\} \tag{5.18}$$

Vicinity aspect (A_k) at the point p_k given by Equation 5.19, is the ratio of the height h_k to the width w_k of the bounding box enclosing the vicinity points $\{p_{k-1}, p_k, p_{k-2}\}$ as shown in Figure 5.4.

$$A_k = \frac{h_k}{w_k} \tag{5.19}$$

The ten statistical features as in Equation 5.9 were computed from the vicinity aspect features to form the aspect features set (see Equation 5.20). Where $aspect_{k,j}$ is the aspect feature set index over k frames and j videos, and $\mathcal{S}^{A_{k,j}}$ is the ten statistical features derived from the computed vicinity aspect.

$$aspect_{k,j} = \{\mathcal{S}^{A_{k,j}}\} \tag{5.20}$$

Vicinity slope (SL_k) as in Equation 5.21 is computed as the cosine of the angle α_k of the straight line from the first to last vicinity point (see Figure 5.4).

$$SL_k = \arccos\left(\frac{(x_{k+1} - x_{k-1})}{\sqrt{(x_{k+1} - x_{k-1})^2 + (y_{k+1} - y_{k-1})^2}}\right) \tag{5.21}$$

The ten statistical features as in Equation 5.9 were computed from the vicinity slope features to form this features set (see Equation 5.22). Where $slope_{k,j}$ is the vicinity slope feature set index over k frames and j videos, and $\mathcal{S}^{SL_{k,j}}$ is the ten statistical features derived from the computed vicinity slope.

$$slope_{k,j} = \{\mathcal{S}^{SL_{k,j}}\} \tag{5.22}$$

Vicinity linearity (LN_k) is the average square distance between all the points within the vicinity and the line joining the first and the last point in the vicinity. For the three point vicinity, there is only one such distance d_k (see Figure 5.4). Therefore, in this case, linearity can be defined as $LN_k = {d_k}^2$. Where d_k is the perpendicular distance from the point p_k and to the line $|p_{k-1}pk+1|$. Again, the ten statistical features as in Equation 5.9 were computed from the vicinity linearity features to form vicinity linearity features set

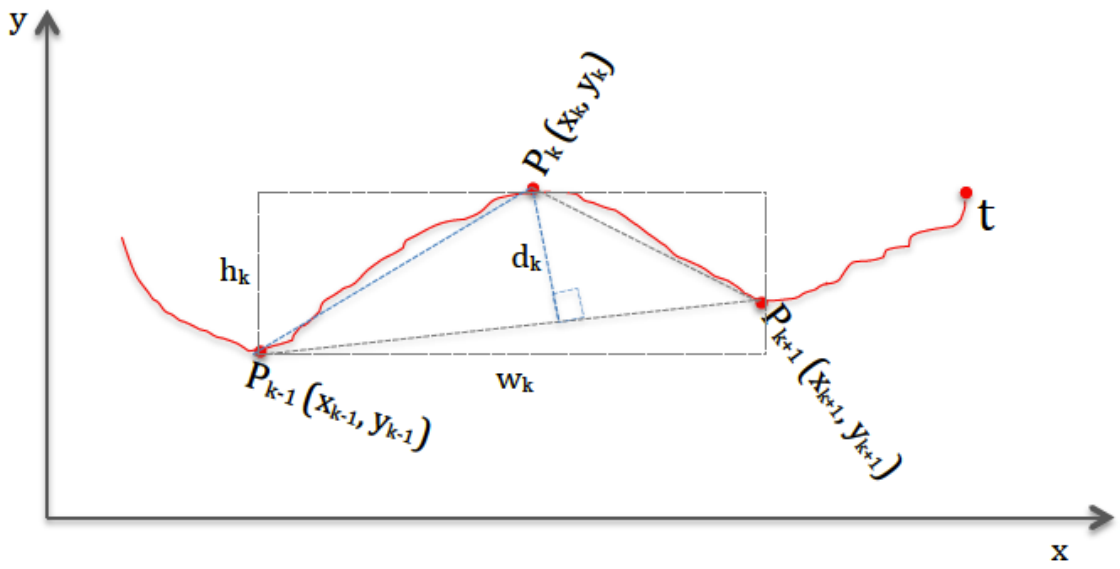


Figure 5.4: Three points vicinity, with a bounding box enclosed. The height (h_k) and width (w_k) of the bounding box are measured for the computation of vicinity aspect. The vicinity slope is measured as the cosine of the angle between the first (p_{k-1}) and last (p_{k+1}) point in the vicinity. Vicinity curliness is measured as total length of the trajectory within the vicinity (that is the length from p_{k-1} through p_k and to p_{k+1}) divided by the maximum of the changes in x and y within the vicinity. Vicinity linearity is the average square distance (d_k) from point p_k to the line joining points p_{k-1} and p_{k+1}.

(see Equation 5.23). Where $linearity_{k,j}$ is the vicinity linearity feature set index over k frames and j videos, and $\mathcal{S}^{SL_{k,j}}$ is the ten statistical features derived from the computed vicinity linearity.

$$linearity_{k,j} = \{\mathcal{S}^{LN_{k,j}}\} \tag{5.23}$$

The feature sets from vicinity curliness, aspect, slope and linearity were then concatenated to form a 40-feature set for vicinity (see Equation 5.24), which have been used to classify species.

$$vicinity_{k,j} = \{curliness_{k,j}, aspect_{k,j}, slope_{k,j}, linearity_{k,j}\} \tag{5.24}$$

Where *vicinity* is the sequence of vicinity features used as part of the motion feature set, index over k frames and j videos.

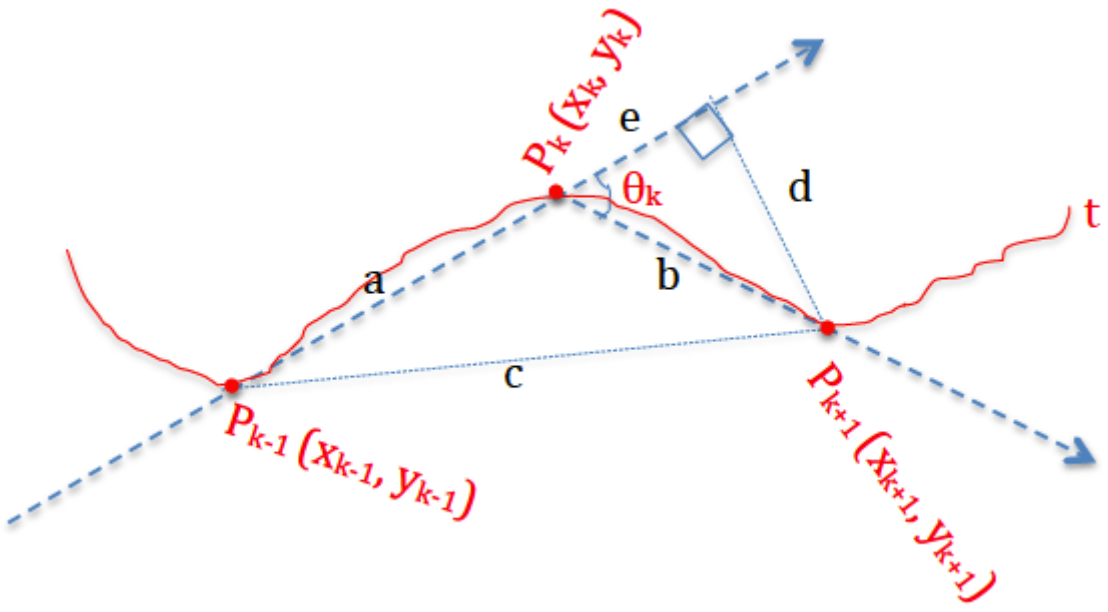


Figure 5.5: The figure shows how the curvature θ_k is measured given three points, the predecessor P_{k-1} and successor P_{k+1} points of the point P_k.

5.2.6 Curvature

Liwicki et al. (2006), Chen et al. (2010) and Graves et al. (2009) have used curvature features successfully in handwriting recognition to determine the writing direction and have achieved remarkable results. The only work that have applied this directly to animal trajectories is that by Beyan (2015), which achieved good results as part of a feature set. Even though this has not yet been used for bird trajectories, bird flight has directional bearings, which can be measured between frames.

Curvature features have been used in this thesis to represent the shape of bird trajectories. This is computed as the cosine of angle between the lines from a point to its predecessor point and the one after. From figure 5.5, with trajectory t_k and three successive points represented by $P_{k-1}, P_k and P_{k+1}$, the curvature $\cos(\theta_k)$ is given by the cosine of angle between the lines $|P_{k-1}P_k|$ and $||P_kP_{k+1}|$ as shown by Equation 5.26.

$$\cos(\theta_k) = \frac{c^2 - a^2 - b^2}{2ab} \tag{5.25}$$

Where: a is the distance from trajectory point $P_{k-1}(x_{k-1}, y_{k-1})$ to $P_k(x_k, y_k)$ given by $\sqrt{(x_{k-1} - x_k)^2 + (y_{k-1} - y_k)^2}$. Similarly, b is given by $\sqrt{(x_k - x_{k+1})^2 + (y_k - y_{k+1})^2}$ and

c is given by $\sqrt{(x_{k-1}-x_{k+1})^2+(y_{k-1}-y_{k+1})^2}$. $k=2\ldots Q-1$ and $Q=64$ is the length of the short trajectory defined in Section 5.1

The curvature values were used to calculate statistical moments to represent curvature features. In total ten features including mean, maximum, minimum, standard deviation, number of zero crossings, local minima and maxima, skewness, energy and entropy of the curvature $\cos(\theta_k)$ were extracted (see Equation 5.26).

$$cur_{k,j} = \{\theta^{\mathcal{S}}_{k,j}\} \tag{5.26}$$

Where $\cos(\theta)^{\mathcal{S}}$ is the ten statistics derived from $\cos(\theta)$ in Equation 5.25 over $k=2\ldots Q-1$ and j the video number and *cur* is the sequence of curvature features used as part of the motion feature set also over $k=2\ldots Q-1$ and j the video number.

The motion features extracted included curvature scale space, turn based, centroid distance function, vicinity and curvature based on sine and cosine. Wing-beat frequency features were also extracted from the width, height and diagonals of the rotated bounding box fitted on the bird's silhouette. For some of the motion features, the mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings were computed. All these features have been concatenated to form the proposed motion feature set (see Equation 5.27) and used to classify bird species.

$$F_{k,j} = \{fft_{k,j}, css_{k,j}, cdf_{k,j}, trn_{k,j}, vic_{k,j}, cur_{k,j}\} \tag{5.27}$$

Where fft is the 27 wing beat frequency features extracted, *css* is the 22 curvature scale space features, cdf is the 10 centroid distance function, trn is the 62 turn based features, *vic* is the 20 vicinity features and *cur* is the 10 curvature features. All these features are computed over k frames of the trajectory and j videos in the dataset.

5.3 Experiments

Two sets of experimental evaluations were performed which have been presented in this chapter. They include:

- Quantifying the effectiveness of the motion feature set, across the dataset of thirteen classes (eleven bird species, one with three colour forms), using four different classifiers.

- Quantifying the effectiveness of the complete feature set (combined appearance and motion), across the dataset of thirteen classes (eleven bird species, one with three colour forms), using the four different classifiers. This was compared with previous results in Chapter 4.

The tools used for these experiments, the setup and classifier settings were the same as reported in Chapter 4. In total, there were 320 features used, which consist of 169 appearances and 151 motion features. 5-fold cross-validation process was used for validation as discussed in Chapter 4. The standard classifiers settings were also the same as explained in Chapter 4.

5.4 Results

Two sets of results were presented and evaluated in this section. The first result set uses only motion features to classify species and the second a combination of appearance and motion features. These were evaluated on dataset #2 to determine the one that increases the correct classification rates and improve differentiation of species with fine-grained appearances.

5.4.1 Evaluation of the Motion Feature Set Results

Table 5.2 shows the results (correct classification rate plus/minus the standard deviation across the five folds) of the motion feature set using the expanded set of thirteen classes (eleven bird species, one with three colour forms), with the four standard classifiers. Table 5.3 presents the confusion matrix, which uses the RF classifier since it is the best

performing classifier. The cross-species confusion matrices for the other classifiers have been presented in Appendix B.

Firstly, the motion features results in Table 5.2 shows the RF, SVM and NB classifiers to outperformed the RT by between 7% - 13%. The Random Forest (RF) classifier gives the highest correct classification rate on the extended dataset, thus outperforming the NB, SVM and RT classifiers by 7%, 8% and 13% respectively. To confirm that the Random Forest (RF) classifier's correct classification rate is statistically significant compared with the other classifiers, the Wilcoxon's test statistic W is calculated. For RF compared with NB classifiers, $W = 0$ (the smaller of $W^+ = 717$ and $W^- = 0$), when RF and RT are compared, $W = 0$ (this is the smaller of $W^+ = 670$ and $W^- = 0$) and $W = 0$ (the smaller of $W^+ = 668.5$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF classifier resulted in a W less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB classifier. This is also true for the RF compared with the RT and SVM classifiers, as both resulted in a W which is less than 415. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the seven species dataset is used with the appearance feature sets.

Secondly, using motion cues alone resulted in a significant decrease in the classification performance when compared with the proposed appearance features as well as Marini et al.'s feature set. Considering the extended dataset with the motion cues alone, the correct classification rates for all classifiers reduced by between 29-47% when compared with the proposed appearance features and between 23-39% when compared with Marini et al.'s. This is an indication that using motion cues alone may not efficiently discriminate species.

Appearance-Related Species

Thirdly, an observation of whether motion cues alone can help differentiate between species of similar appearance was investigated. Given that the RF classifier has shown

Table 5.2: Correct classification rates by species, based on **the motion features** without feature selection using the four standard classifiers. These results use the **thirteen classes dataset** introduced in the previous chapter and shows ± standard deviation across the five folds.

Species	**NB**	**RF**	**RT**	**SVM**
a=Alexandrine Parakeet	3% ± 0.14%	21% ± 0.82%	13% ± 0.66%	21% ± 0.76%
b=Nanday Parakeet	7% ± 0.34%	15% ± 0.14%	10% ± 0.37%	11% ± 0.47%
c=Blue-crowned Parakeet	10% ± 0.03%	9% ± 0.37%	9% ± 0.03%	12% ± 0.46%
d=Common House Martin	13% ± 0.08%	45% ± 0.04%	26% ± 0.52%	37% ± 0.09%
e=Eastern Rosella	19% ± 0.60%	9% ± 0.62%	5% ± 0.11%	6% ± 0.12%
f=Budgerigar (yellow)	11% ± 0.64%	1% ± 0.00%	7% ± 0.09%	3% ± 0.89%
g=House Sparrow	5% ± 0.24%	15% ± 0.52%	11% ± 0.53%	9% ± 0.25%
h=Budgerigar (wild-type)	1% ± 0.00%	1% ± 0.00%	6% ± 0.51%	6% ± 0.48%
i=Common Wood Pigeon	17% ± 0.80%	2% ± 0.77%	5% ± 0.87%	5% ± 0.73%
j=Black-headed Gull	85% ± 0.47%	85% ± 0.03%	53% ± 0.13%	68% ± 0.83%
k=Cockatiel	25% ± 0.12%	5% ± 0.31%	10% ± 0.74%	7% ± 0.36%
l=Budgerigar (blue)	6% ± 0.12%	7% ± 0.38%	12% ± 0.53%	3% ± 0.45%
m=Common Starling	13% ± 0.03%	24% ± 0.18%	15% ± 0.29%	46% ± 0.21%
Overall Correctly Classified	**32% ± 0.38%**	**38% ± 0.31%**	**25% ± 0.47%**	**33% ± 0.58%**

Table 5.3: The confusion matrix based on the **Random Forest** classifier without feature selection, using **the motion features** on the **thirteen classes dataset (eleven bird species, one with three colour forms)**. %CC is the percentage corectly classified.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	21.1%	9.7%	6.5%	25.2%	0.6%	1.1%	2.7%	0.4%	0.3%	22.5%	3.1%	2.1%	4.7%	21%	7845
b=Nanday Parakeet	19.1%	15.4%	7.9%	22.9%	0.8%	1.1%	2.6%	0.3%	0.2%	17.2%	3.9%	2.0%	6.7%	15%	6246
c=Blue-crowned Parakeet	20.1%	12.8%	9.3%	24.3%	0.8%	0.9%	3.1%	0.6%	0.2%	15.5%	3.5%	1.8%	7.2%	9%	5332
d=Common House Martin	4.6%	2.5%	1.4%	45.3%	0.3%	0.6%	1.0%	0.2%	0.0%	40.9%	0.9%	1.6%	0.8%	45%	17896
e=Eastern Rosella	12.9%	8.0%	5.7%	21.0%	9.4%	0.7%	1.9%	0.4%	0.3%	30.8%	2.8%	1.9%	4.2%	9%	3247
f=Budgerigar (yellow)	8.2%	4.6%	2.8%	10.1%	0.3%	1.3%	2.2%	23.5%	0.1%	17.6%	1.4%	26.4%	1.5%	1%	4329
g=House Sparrow	11.2%	6.6%	4.1%	24.2%	0.4%	0.9%	14.8%	0.6%	0.2%	27.4%	1.4%	1.8%	6.4%	15%	5318
h=Budgerigar (wild-type)	9.5%	5.9%	3.5%	11.3%	0.4%	16.5%	4.2%	1.2%	0.2%	10.2%	1.9%	32.1%	3.1%	1%	3349
i=Common Wood Pigeon	21.5%	12.4%	8.5%	21.8%	0.8%	1.0%	3.5%	0.3%	2.3%	16.8%	3.7%	1.9%	5.4%	2%	2027
j=Black-headed Gull	1.5%	0.7%	0.4%	11.2%	0.1%	0.2%	0.2%	0.1%	0.0%	84.9%	0.3%	0.4%	0.1%	85%	29695
k=Cockatiel	13.7%	9.1%	5.0%	25.0%	0.6%	0.6%	1.4%	0.3%	0.2%	33.5%	5.3%	1.7%	3.6%	5%	5687
l=Budgerigar (blue)	6.2%	3.4%	2.1%	18.4%	0.4%	15.4%	2.1%	14.5%	0.2%	28.1%	1.1%	6.8%	1.2%	7%	7030
m=Common Starling	17.2%	12.5%	6.9%	16.8%	0.5%	0.8%	5.4%	0.4%	0.3%	10.9%	2.9%	1.3%	24.0%	24%	5392
Overall Correctly Classified														**38%**	

best overall performance, a detailed cross-species confusion matrix for this classifier, obtained using motion features with the larger dataset, shown in Table 5.3 was provided. Also, Figure 5.6 shows examples of the Alexandrine Parakeet, Nanday Parakeet and Blue-crowned Parakeet, which have closely related appearances. However, examples of their flight trajectories appear dissimilar (Figure 5.7). Now, when motion features alone were used to classify these species, more misclassifications were recorded than when appearance features alone were used. For example, 2.8% more Alexandrine Parakeets misclassified as Nanday Parakeets; 12% more Nanday Parakeets as Alexandrine Parakeets; and 12% more Blue-crowned Parakeets as Alexandrine Parakeets when motion features were used alone compared with appearance. This suggests that using motion cues alone may not be sufficient to differentiate between species with closely-related appearances but when used as a weak classifier may differentiate these species. Even through motion cues were a weak classifier, the classification rates suggests that it is able to differentiate between species with similar appearance than when only appearance features were used.

Motion-Related Species

There are three Budgerigar species with nearly the same flight pattern but different colour forms and were mainly misclassified as other Budgerigar. In particular, 24% and 26% Budgerigar (yellow) were misclassified as Budgerigar(wild-type) and Budgerigar (blue) respectively, 17% and 32% Budgerigar (wild-type) were misclassified as Budgerigar (yellow) and Budgerigar (blue) respectively, and 16% and 15% Budgerigar (blue) were misclassifies as Budgerigar (yellow) and Budgerigar (wild-type) respectively. This means that half the Budgerigar (yellow) and Budgerigar(wild-type) were misclassified as other Budgerigars and more than a third of the Budgerigar (blue) were misclassified as other Budgerigars. This is mainly because Budgerigar species only vary by colour and not motion, and thus explaining why their correct classification rates are low when motion features are used.

Figure 5.6: Fine-grained species (Species closely related by appearances) - From left to right is Alexandrine Parakeet (*Psittacula eupatria*), b) Nanday Parakeet (*Aratinga nenday*), c) Blue-crowned Parakeet (*Thectocercus acuticaudatus*)

Figure 5.7: Sample tracks of fine-grained species (that are species closely related by appearances) - From left to right is Alexandrine Parakeet (*Psittacula eupatria*), b) Nanday Parakeet (*Aratinga nenday*), c) Blue-crowned Parakeet (*Thectocercus acuticaudatus*)

Majority and Minority Sampled Species

Since the datasets were imbalanced, the minority class is expected to produce a low classification rate (Ramyachitr and Manikandan, 2014; Lee et al., 2009). Most machine learning algorithms sometimes fail to classify imbalanced data because the classification error in the majority class is dominated by that of the minority (Ramyachitr and Manikandan, 2014). The RF classifier shows 85% correct classification rate for the majority class using the thirteen classes dataset. This is 83% higher than the minority class, which is only 2%. Thus, demonstrating the effect of this imbalanced dataset (imbalanced ratio of 10.1) when motion features are used and the likelihood of redundancy with this feature. The imbalanced in the dataset also explains why some species with small samples were generally misclassified as species with larger samples (Black-headed Gulls and Common House Martin). In particular, 21% and 31% of Eastern Rosella, 18% and 28% of Budgerigar (blue), 10% and 18% Budgerigar (yellow) and 12% and 11% Budgerigar (wild-type) were misclassified as Black-headed Gulls and Common House Martin respectively. This

Table 5.4: Table showing the number of videos and images for four randomly undersampled datasets based on the thirteen classes dataset. RUS-79 randomly undersampled the dataset so that all classes with videos greater than 79 were reduced to 79 videos, RUS-59 randomly undersampled classes to 59 videos, RUS-44 to 44 videos and RUS-37 to 37 videos.

Species	RUS-79		RUS-59		RUS-44		RUS-37	
	# of videos	# of images	# of videos	# of images	# of videos	# of images	# of videos	# of images
a=Alexandrine Parakeet	79	7845	59	5618	44	3840	37	3041
b=Nanday Parakeet	60	6246	59	6026	44	4450	37	3999
c=Blue-crowned Parakeet	60	5332	59	5307	44	4343	37	3320
d=Common House Martin	79	8970	59	7505	44	5998	37	4771
e=Eastern Rosella	44	3247	44	3247	44	3247	37	2378
f=Budgerigar (yellow)	54	4329	54	4329	44	3147	37	2876
g=House Sparrow	78	5318	59	4098	44	2871	37	2427
h=Budgerigar (wild-type)	48	3349	48	3349	44	3268	37	2790
i=Common Wood Pigeon	37	2027	37	2027	37	2027	37	2027
j=Black-headed Gull	79	15065	59	13069	44	10364	37	8973
k=Cockatiel	59	5687	59	5687	44	4630	37	4180
l=Budgerigar (blue)	79	7030	59	5199	44	4093	37	3111
m=Common Starling	71	5392	59	4496	44	3327	37	2662
Totals	**827**	**79837**	**714**	**69957**	**565**	**55605**	**481**	**46555**

demonstrates that not only were Budgerigar misclassified among each other as explained in the previous paragraph, but they were also misclassified as species with larger samples.

Random Under-Sampling (RUS) of the Extended Dataset

The objective of this section is to apply the RUS imbalanced dataset techniques to the extended dataset and using the motion features to determine if misclassification of samples in the minority classes as the majority classes are reduced.

Four datasets were formed from the extended dataset by randomly undersampling the videos in each class in an attempt to balance the dataset. Table 5.4 shows the number of videos and images in each undersampled dataset. The first dataset (RUS-79) was obtained by randomly undersampling classes with more than 79 videos to 79, the second dataset (RUS-59) was obtained by randomly undersampling to 59 videos, and the third (RUS-44) was randomly undersampled to 44 videos. The final dataset, RUS-37, was obtained by randomly undersampling all the other classed to the size of the minority class (37 videos). Species were then classified using the best performing classifier (random forest) and the motion features extracted from these datasets.

Table 5.6 presents confusion matrix of classification using the RF classifier and RUS-37 dataset. This is the more balanced dataset, which have been used to explore the effect of balancing the dataset on smaller sampled classes misclassified as larger ones. Table

Table 5.5: Correct classification rates by species, based on **the motion features** without feature selection using the RF classifier. These results use the **thirteen classes dataset** introduced in the previous chapter but resampled using Random Undersampling (RUS). RUS >79 randomly undersampled the dataset so that all classes with videos greater than 79 were reduced to a maximum of 79 videos, RUS >59 randomly undersampled classes to a maximum of 59 videos and RUS >44 to a maximum of 44 videos. All results shows ± standard deviation across the five folds.

Species	**RUS-79**	**RUS-59**	**RUS-44**	**RUS-37**
a=Alexandrine Parakeet	29% ± 0.21%	18% ± 0.71%	13% ± 0.38%	12% ± 0.78%
b=Nanday Parakeet	20% ± 0.31%	24% ± 0.39%	23% ± 0.44%	27% ± 0.54%
c=Blue-crowned Parakeet	12% ± 0.26%	16% ± 0.43%	19% ± 0.69%	16% ± 0.71%
d=Common House Martin	29% ± 0.11%	27% ± 0.99%	27% ± 0.14%	27% ± 0.28%
e=Eastern Rosella	3% ± 0.61%	3% ± 0.79%	12% ± 0.56%	12% ± 0.28%
f=Budgerigar (yellow)	4% ± 0.87%	8% ± 0.49%	9% ± 0.77%	9% ± 0.29%
g=House Sparrow	19% ± 0.27%	18% ± 0.75%	14% ± 0.27%	16% ± 0.61%
h=Budgerigar (wild-type)	2% ± 0.01%	3% ± 0.00%	9% ± 0.39%	10% ± 0.43%
i=Common Wood Pigeon	3% ± 0.11%	4% ± 0.13%	14% ± 0.26%	17% ± 0.72%
j=Black-headed Gull	81% ± 0.07%	82% ± 0.48%	82% ± 0.24%	82% ± 0.46%
k=Cockatiel	10% ± 0.15%	14% ± 0.45%	14% ± 0.74%	17% ± 0.01%
l=Budgerigar (blue)	15% ± 0.65%	13% ± 0.23%	13% ± 0.54%	14% ± 0.53%
m=Common Starling	29% ± 0.12%	24% ± 0.21%	19% ± 0.22%	17% ± 0.03%
Overall Correctly Classified	**29% ± 0.34%**	**29% ± 0.41%**	**29% ± 0.49%**	**30% ± 0.56%**

5.5 shows the correct classification rates ± standard deviation across the five folds for the four randomly undersampled dataset.

The result of RUS-37 dataset shows 82% correct classification rate for the majority class (3% less than when the entire dataset was used). This is 65% more than the minority class, which is 17%. The correct classification rate of the minority class increased (RUS-79 is 3%, RUS-59 is 4%, RUS-44 is 14% and RUS-37 is 17%) as the difference in samples between the majority and minority classes were reduced. This means that balancing the classes may reduce the bias in classification rates among classes. However, there was an unnoticeable or no reduction in the overall classification rates as the number of samples are reduced from RUS-79 to RUS-37.

Even though the species with smaller samples (Budgerigar classes) were still misclassified as those with larger samples (particularly, Black-headed Gulls and Common House Martin), the misclassification reduced (see Table 5.5) when the RUS-37 is compared with the dataset without RUS. In particular, 4.3% and 13.6% less Eastern Rosella, 5.8% and 7.2% less Budgerigar (blue), 8.7% and 2.3% less Budgerigar (yellow) and 0.8%

Table 5.6: The confusion matrix based on the **Random Forest** classifier without feature selection, using **the motion features** on the **thirteen classes dataset**. The dataset have been randomly undersampled so that the number of video in classes with videos more than 37 are reduced to a maximum of 37 videos. %CC is the percentage corectly classified.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC
a=Alexandrine Parakeet	12.0%	16.5%	8.8%	8.5%	2.2%	4.3%	2.5%	2.6%	1.9%	24.5%	9.7%	2.7%	3.7%	12%
b=Nanday Parakeet	8.8%	27.1%	11.8%	6.5%	2.4%	3.3%	2.5%	2.3%	2.4%	16.1%	10.6%	2.1%	4.3%	27%
c=Blue-crowned Parakeet	9.9%	20.7%	16.0%	5.7%	2.7%	3.6%	3.3%	3.3%	3.0%	13.2%	11.2%	2.8%	4.5%	16%
d=Common House Martin	3.5%	6.1%	3.2%	27.3%	1.3%	3.3%	1.4%	2.1%	0.7%	44.6%	3.3%	2.3%	0.9%	27%
e=Eastern Rosella	7.8%	14.1%	8.3%	7.4%	11.8%	3.2%	1.9%	3.0%	1.9%	26.5%	9.1%	2.4%	2.5%	12%
f=Budgerigar (yellow)	8.0%	11.1%	9.8%	7.8%	2.0%	9.3%	3.3%	15.5%	1.3%	8.9%	7.0%	13.8%	2.2%	9%
g=House Sparrow	7.3%	15.6%	8.5%	8.0%	1.9%	4.9%	16.2%	4.5%	2.1%	17.5%	6.5%	2.9%	4.1%	16%
h=Budgerigar (wild-type)	6.2%	12.1%	7.0%	9.2%	2.3%	16.0%	3.4%	10.0%	3.4%	9.4%	7.6%	11.1%	2.3%	10%
i=Common Wood Pigeon	11.0%	13.7%	14.2%	4.8%	2.8%	3.6%	3.1%	3.1%	16.5%	11.3%	10.5%	2.2%	3.5%	17%
j=Black-headed Gull	1.7%	2.3%	1.1%	6.9%	0.6%	1.0%	0.3%	0.9%	0.3%	82.4%	1.6%	0.7%	0.1%	82%
k=Cockatiel	8.8%	16.8%	8.1%	7.6%	2.3%	3.3%	1.7%	2.4%	1.5%	25.0%	17.4%	2.0%	3.1%	17%
l=Budgerigar (blue)	7.0%	8.6%	6.2%	11.2%	1.8%	7.7%	3.3%	9.0%	1.4%	22.3%	5.9%	13.6%	2.2%	14%
m=Common Starling	8.2%	19.7%	10.3%	6.1%	1.7%	3.0%	3.6%	2.3%	2.4%	12.8%	11.2%	1.8%	16.9%	17%
Overall Correctly Classified														**30%**

and 2.1% less Budgerigar (wild-type) were misclassified as Black-headed Gulls and Common House Martin respectively. This demonstrates that balancing the dataset has reduced misclassification of samples in the minority classes as the majority ones.

5.4.2 Evaluation of the Full Feature Set Results

It has been demonstrated in the previous chapter that the appearance features outperformed those of Marini et al. on both small and extended datasets using all four standard classifiers (Random Forest, Random Tree, Naive Bayes and SVM classifiers). This section presents an expanded set of comparative results which incorporate the full feature set of combined features based on the expanded video dataset of thirteen classes (eleven bird species, one with three different colour forms).

Table 5.7 shows the results (that is correct classification rate plus/minus the standard deviation across the five folds) of the expanded set of thirteen classes (eleven bird species, one with three colour forms).

Firstly, the combined appearance and motion features results in Table 5.7, which shows that the RF, SVM and NB classifiers outperformed the RT by between 4% - 19%. Again, the Random Forest (RF) classifier gives the highest correct classification rate with all combinations of features, thus the assertion that this is the most effective classifier of the four tested for this problem domain. There was evident with the extended dataset, where the RF classifier outperformed NB, SVM and RT classifiers by 15%, 8% and 19%

Table 5.7: Correct classification rates by species, based on **the combined features** without feature selection using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in Chapter 4 and shows ± standard deviation across the five folds.

Species	**NB**	**RF**	**RT**	**SVM**
a=Alexandrine Parakeet	40% ± 0.62%	81% ± 0.84%	52% ± 0.07%	60% ± 0.55%
b=Nanday Parakeet	53% ± 0.88%	80% ± 0.13%	50% ± 0.52%	56% ± 0.77%
c=Blue-crowned Parakeet	57% ± 0.75%	83% ± 0.21%	59% ± 0.43%	66% ± 0.35%
d=Common House Martin	90% ± 0.06%	99% ± 0.67%	84% ± 0.51%	98% ± 0.35%
e=Eastern Rosella	51% ± 0.32%	70% ± 0.77%	44% ± 0.63%	44% ± 0.15%
f=Budgerigar (yellow)	70% ± 0.13%	79% ± 0.52%	51% ± 0.32%	78% ± 0.03%
g=House Sparrow	59% ± 0.53%	63% ± 0.04%	44% ± 0.19%	52% ± 0.83%
h=Budgerigar (wild-type)	34% ± 0.57%	45% ± 0.61%	33% ± 0.05%	33% ± 0.49%
i=Common Wood Pigeon	55% ± 0.65%	53% ± 0.71%	39% ± 0.25%	48% ± 0.25%
j=Black-headed Gull	84% ± 0.14%	99% ± 0.11%	85% ± 0.17%	97% ± 0.77%
k=Cockatiel	63% ± 0.13%	81% ± 0.25%	49% ± 0.64%	68% ± 0.36%
l=Budgerigar (blue)	56% ± 0.56%	68% ± 0.74%	49% ± 0.68%	53% ± 0.89%
m=Common Starling	70% ± 0.37%	71% ± 0.46%	45% ± 0.81%	75% ± 0.37%
Overall Correctly Classified	**69% ± 0.44%**	**85% ± 0.47%**	**65% ± 0.41%**	**76% ± 0.47%**

respectively. The SVM classifier similarly had a better performance than either NB or Random Tree on the same dataset. Observing cross-species correct classification rates of the RF and SVM classifiers show that the RF classifier outperformed the SVM classifier on all species except the Common Starling class, where the SVM had 10% more correct classification rate than the RF.

Again, the statistical significance of the Random Forest (RF) classifier's correct classification rate compared with the other classifiers were undertaken by computing the Wilcoxon's test statistic W. When the RF classifier was compared with NB, $W = 0$ (the smaller of $W^+ = 665$ and $W^- = 0$), $W = 0$ (this is the smaller of $W^+ = 577.5$ and $W^- = 0$) when RF and RT were compared and $W = 0$ (the smaller of $W^+ = 640.5$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF resulted in a $W = 0$, which is less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB. This is also true for the RF compared with the RT and SVM classifiers. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the thirteen classes dataset is used with the appearance feature sets.

Comparing the combined feature set (motion and appearance) with only the appearance features (see Chapter 4), it is evident that the naive addition of the motion features makes little difference to classifier performance. Considering the combined feature set, the addition of motion features increased the performance of the NB and SVM classifiers by 15% and 6% respectively and decreased the performance of RF and RT classifiers by 1% and 3% respectively. The results in Section 5.4.1 showed that motion features can be used as a weak classifier, motivating the exploration of effective feature selection, which will be presented in the next section.

Appearance-Related Species

This is further motivated by an examination of the fine-grained performance of the proposed method; that is, differentiation between species of similar appearance. Given that the RF classifier has shown best overall performance, a detailed cross-species confusion matrix for this classifier was provided which was obtained using combined features with the larger dataset, shown in Table 5.8. The cross-species confusion matrix for the other classifiers have been presented in Appendix C. When appearance features alone are used to classify these species, more misclassifications are recorded than when combined (appearance and motion features) are used. For example, the misclassification of Alexandrine Parakeets as Nanday Parakeets reduce by 0.4%; Nanday Parakeets as Alexandrine Parakeets by 0.8%; and Blue-crowned Parakeets as Alexandrine Parakeets by 1.9% the combined results are compared with appearance. This suggests that merging appearance and motion features may increase the correct classification rates for species with closely-related appearances.

Motion-Related Species

The three Budgerigar species have different colour form, thus the low misclassification among these species when only appearance features were used in Chapter 4. In this Chapter, when motion features alone are used for classification, nearly half of Budgerigar(yellow) and Budgerigar(wild-type) were misclassified as other Budgerigars. Using the combined features, misclassification among these species reduced compared to when

Table 5.8: The confusion matrix based on the **Random Forest** classifier without feature selection, using **the combined features** on the **thirteen classes dataset (eleven bird species, one with three colour forms)**. %CC is the percentage corectly classified.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	80.6%	6.6%	2.1%	0.5%	0.2%	1.3%	0.5%	0.8%	0.2%	3.1%	1.4%	2.4%	0.4%	81%	7845
b=Nanday Parakeet	6.4%	79.6%	3.5%	3.4%	0.4%	1.8%	0.2%	0.4%	0.6%	0.8%	1.2%	0.8%	0.8%	80%	6246
c=Blue-crowned Parakeet	6.3%	4.9%	82.6%	1.5%	0.4%	0.8%	0.3%	0.6%	0.1%	0.6%	0.3%	1.4%	0.2%	83%	5332
d=Common House Martin	0.1%	0.3%	0.0%	98.8%	0.0%	0.0%	0.2%	0.0%	0.0%	0.4%	0.1%	0.0%	0.1%	99%	17896
e=Eastern Rosella	3.4%	3.5%	2.1%	1.3%	69.8%	1.7%	0.8%	0.9%	0.2%	8.6%	2.2%	5.3%	0.3%	70%	3247
f=Budgerigar (yellow)	6.7%	3.3%	1.5%	1.0%	0.4%	79.0%	1.4%	1.2%	0.0%	2.6%	1.5%	1.1%	0.2%	79%	4329
g=House Sparrow	1.0%	0.2%	0.2%	18.2%	0.2%	0.9%	62.6%	0.3%	0.0%	5.4%	1.2%	1.2%	8.6%	63%	5318
h=Budgerigar (wild-type)	12.1%	6.4%	5.5%	2.7%	1.5%	6.3%	2.6%	44.9%	0.5%	9.6%	3.6%	3.5%	0.9%	45%	3349
i=Common Wood Pigeon	4.8%	8.8%	5.2%	3.5%	0.5%	0.6%	1.0%	0.4%	53.2%	12.9%	3.8%	1.5%	3.7%	53%	2027
j=Black-headed Gull	0.1%	0.1%	0.0%	0.5%	0.1%	0.0%	0.1%	0.0%	0.0%	98.8%	0.1%	0.3%	0.0%	99%	29695
k=Cockatiel	0.4%	0.5%	0.1%	5.9%	0.4%	0.6%	1.4%	0.2%	0.0%	5.5%	80.5%	0.4%	4.2%	81%	5687
l=Budgerigar (blue)	2.8%	0.9%	0.3%	2.1%	0.7%	0.6%	0.7%	0.6%	0.1%	21.4%	1.8%	67.6%	0.4%	68%	7030
m=Common Starling	1.1%	0.6%	0.3%	10.9%	0.1%	0.9%	10.3%	0.1%	0.0%	1.5%	3.1%	0.3%	70.8%	71%	5392
Overall Correctly Classified														**85%**	

motion features were used. Nearly one-twentieth of these Budgerigar species were misclassified as other Budgerigars. Particularly, 1.2% and 1.1% Budgerigar (yellow) were misclassified as Budgerigar(wild-type) and Budgerigar (blue) respectively, 6.3% and 3.5% Budgerigar (wild-type) were misclassified as Budgerigar (yellow) and Budgerigar (blue) respectively, and 0.6% Budgerigar (blue) were misclassified as Budgerigar (yellow) and Budgerigar (wild-type) respectively. This suggests that even though motion features causes misclassification among these species, their distinct colour help reduce this when combined features are used.

Majority and Minority Sampled Species

On the thirteen classes dataset, the classification rate for the majority class (Black-headed Gull) is between 85% - 99% for all four standard classifiers and the minority class (Common Wood Pigeon) is between 39% - 55% (see Table 5.7). The minority class was not the class with the lowest classification rate but rather the Budgerigar (wild-type) class, with a classification rate of approximately 18% - 45% across all four classifiers. The confusion matrix Table 5.8 shows that 12.9% of minority class (Common Wood Pigeon) were misclassified as majority class (Black-headed Gull). This was 9.6% for the Budgerigar (wild-type) class, which was the next class with lower samples after the Common Wood Pigeon class. However, Budgerigar (wild-type) had the lowest correct classification rate as nearly a quarter (24.1%) were misclassified as Parakeets (12.1% as Alexandrine Parakeet, 6.4% as Nanday Parakeet and 5.6% as Blue-crowned Parakeet). The reason for this misclassification is that Budgerigar (wild-type) is more related in appearance (colour) to

Table 5.9: Correct classification rates by species, based on **the combined features** without feature selection using the RF classifier. These results use the **thirteen classes dataset** introduced in the previous chapter but resampled using Random Undersampling (RUS). RUS-79 randomly undersampled the dataset so that all classes with videos greater than 79 were reduced to 79 videos, RUS-59 randomly undersampled classes to 59 videos, RUS-44 to 44 videos and RUS-37 to 37 videos. All results shows ± standard deviation across the five folds.

Species	**RUS-79**	**RUS-59**	**RUS-44**	**RUS-37**
a=Alexandrine Parakeet	81% ± 0.21%	78% ± 0.05%	78% ± 0.58%	82% ± 0.36%
b=Nanday Parakeet	80% ± 0.83%	83% ± 0.31%	80% ± 0.62%	86% ± 0.13%
c=Blue-crowned Parakeet	83% ± 0.29%	86% ± 0.73%	89% ± 0.37%	87% ± 0.14%
d=Common House Martin	97% ± 0.85%	97% ± 0.03%	97% ± 0.84%	97% ± 0.51%
e=Eastern Rosella	72% ± 0.12%	74% ± 0.15%	79% ± 0.65%	80% ± 0.69%
f=Budgerigar (yellow)	81% ± 0.32%	84% ± 0.65%	82% ± 0.19%	84% ± 0.85%
g=House Sparrow	69% ± 0.87%	69% ± 0.56%	66% ± 0.11%	70% ± 0.08%
h=Budgerigar (wild-type)	50% ± 0.78%	54% ± 0.87%	61% ± 0.42%	62% ± 0.02%
i=Common Wood Pigeon	56% ± 0.74%	58% ± 0.64%	63% ± 0.29%	65% ± 0.25%
j=Black-headed Gull	98% ± 0.24%	98% ± 0.59%	98% ± 0.31%	98% ± 0.34%
k=Cockatiel	81% ± 0.54%	82% ± 0.53%	85% ± 0.57%	86% ± 0.16%
l=Budgerigar (blue)	77% ± 0.61%	72% ± 0.11%	72% ± 0.11%	69% ± 0.43%
m=Common Starling	72% ± 0.76%	72% ± 0.72%	70% ± 0.34%	70% ± 0.84%
Overall Correctly Classified	**82% ± 0.48%**	**82% ± 0.56%**	**83% ± 0.57%**	**83% ± 0.49%**

the Parakeet species. The Budgerigar classes were also with smaller samples and therefore misclassified as the majority class (Black-headed Gull). There were 9.6%, 7.2% and 2.7% Budgerigar (wild-type), Budgerigar (blue) and Budgerigar (yellow) respectively, that were misclassified as Black-headed Gull.

Random Under-Sampling (RUS) of the Extended Dataset

The objective of this section is to apply the RUS imbalanced dataset techniques to the extended dataset and using the combined features to determine if misclassification of samples in the minority classes (especially, the three Budgerigar classes) as the majority are reduced. RUS-37 dataset, which was obtained by undersampling videos in classes with more than 37 videos was used. This undersampled dataset has been described in Section 5.4.1 and has been used in this section to perform experiments, as it is the most balanced dataset out of the four undersampled.

Table 5.10 presents confusion matrix of classification using the combined feature set with RUS-37 dataset. This is the more balanced dataset, which was used to explore the effect of balancing the dataset on smaller sampled classes that are being misclassified as

Table 5.10: The confusion matrix based on the **Random Forest** classifier without feature selection, using **the combined features** on the **thirteen classes dataset**. The dataset have been randomly undersampled so that the number of video in classes with videos more than 37 are reduced to a maximum of 37 videos. %CC is the percentage corectly classified.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	81.9%	5.6%	1.7%	0.3%	0.4%	2.8%	0.5%	2.1%	0.6%	1.7%	1.2%	1.1%	0.2%	82%	3041
b=Nanday Parakeet	3.4%	85.8%	2.4%	0.8%	0.5%	1.7%	0.1%	1.0%	1.0%	0.6%	1.6%	0.4%	0.7%	86%	3999
c=Blue-crowned Parakeet	3.9%	3.0%	86.9%	0.2%	0.9%	1.1%	0.2%	1.6%	0.5%	0.5%	0.6%	0.4%	0.2%	87%	3320
d=Common House Martin	0.3%	0.9%	0.1%	96.8%	0.0%	0.0%	0.3%	0.1%	0.1%	0.4%	0.3%	0.0%	0.6%	97%	4771
e=Eastern Rosella	0.9%	2.9%	1.7%	0.3%	80.3%	0.3%	0.3%	1.4%	0.3%	5.7%	3.2%	2.6%	0.0%	80%	2378
f=Budgerigar (yellow)	1.9%	3.9%	1.9%	0.4%	1.3%	83.6%	0.7%	1.9%	0.3%	1.7%	2.2%	0.3%	0.1%	84%	2876
g=House Sparrow	0.4%	0.4%	0.2%	10.3%	0.1%	1.8%	70.3%	0.8%	0.1%	2.4%	2.1%	0.2%	11.0%	70%	2427
h=Budgerigar (wild-type)	5.0%	6.3%	4.0%	1.7%	2.2%	6.3%	1.7%	62.0%	1.5%	4.1%	3.1%	1.5%	0.6%	62%	2790
i=Common Wood Pigeon	2.1%	8.5%	5.4%	2.5%	1.2%	1.5%	1.1%	1.3%	65.3%	1.2%	6.5%	0.5%	2.9%	65%	2027
j=Black-headed Gull	0.2%	0.2%	0.0%	0.6%	0.3%	0.0%	0.1%	0.1%	0.0%	98.0%	0.3%	0.3%	0.0%	98%	8973
k=Cockatiel	0.3%	0.7%	0.1%	3.7%	0.5%	0.6%	1.1%	0.2%	0.1%	3.6%	85.9%	0.7%	2.4%	86%	4180
l=Budgerigar (blue)	0.8%	0.9%	0.2%	1.8%	2.9%	0.6%	0.8%	1.5%	0.1%	18.7%	2.5%	68.7%	0.7%	69%	3111
m=Common Starling	0.3%	1.1%	0.5%	8.4%	0.1%	0.6%	11.1%	0.3%	0.3%	1.2%	6.5%	0.2%	69.6%	70%	2662
Overall Correctly Classified														**83%**	**46555**

larger ones. Table 5.9 shows the correct classification rates $\pm$ standard deviation across the five folds using the RF classifier on the resampled RUS-37, RUS-44, RUS-59 and RUS-79 datasets.

The result of the RUS-37 dataset shows 98% correct classification rate for the majority class (1% less than when the entire dataset was used). The correct classification rate for the minority class is 65% which is 12% higher than when the entire dataset was used (53%). The difference between the majority and minority class reduced by 13%, when the undersampled dataset was used (RUS-37). This may mean that with equal sized classes, the bias in correct classification rate between the majority and minority classes may reduce. Again, there was a noticeable a reduction in the overall classification rates as the number of samples are reduced from RUS-79 to RUS-37.

Although classes with smaller samples (i.e. Budgerigar classes) were still misclassified as those with larger samples, particularly, Black-headed Gulls and Common House Martin, the misclassification reduced when the dataset was undersampled to 37 videos per class (RUS-37). In particular, 2.9% and 1.0% less Eastern Rosella, 2.7% and 0.3% less Budgerigar (blue), 0.9% and 0.6% less Budgerigar (yellow) and 5.4% and 0.1% less Budgerigar (wild-type) were misclassified as Black-headed Gulls and Common House Martin respectively. This again, demonstrates that balancing the dataset may reduce misclassification of species in the minority classes as the majority ones.

Table 5.11: Classifiers training times in seconds.

	Appearance				Motion				Marini			
	NB	SVM	RT	RF	NB	SVM	RT	RF	NB	SVM	RT	RF
Mean	157.51	152.1	99.16	114.21	202.65	202.72	110.12	180.13	85.74	87.45	65.33	97.34
Min	141.95	136.67	89.16	102.32	186.38	186.35	98.24	120.86	66.07	67.66	44.98	71.38
Max	202.33	196.31	108.74	141.22	238.03	238.31	150.56	220.49	157.71	159.94	120.14	170.13
σ	11.00	10.82	6.74	9.97	14.94	15.08	10.01	9.95	19.48	19.64	16.86	21.01

5.4.3 Performance Evaluation

This section empirically compares the computational performance of the proposed classifiers with those of Marini et al. (2013), and ascertain whether they are capable of running in real-time on standard computer hardware. To achieve this the seven species dataset was used to perform experiments on a Mac book pro laptop running OS X 10.9.5, with 2.5 GHz Processor and 4 GB Ram. The algorithms and classifiers were all written in C++ with XCode 5.1.1 and OpenCV 3.0. Both the classification and recognition phases were tested individually.

To compare the performance of the training phase, a software timing function (millisecond accuracy) was utilised by recording the time in millisecond taken to build Random Trees, Random Forest, Naive Bayes and SVM classifiers for the appearance feature set, motion-feature set, and Marini *et al.'s* algorithm. The timings include both image feature extraction and training stages, using a five-fold cross-validation technique (as according to the reported experimental setup). The timings were run for each classifier 50 times so that an estimate a mean value, and also the standard deviation, minimum, and maximum timings in each case will be recorded. Results are shown in Table 5.11.

To compare performance for the classification phase a set of experiments similar to the above were performed. For each classifier, a set of 1500 individual birds were used, and recording the time taken to classify the entire set. This includes both the feature extraction, and the actual classification using Random Trees, Random Forest, Naive Bayes and SVM with the appearance features, motion features, and Marini *et al.'s* feature set. Again, the experimental runs were repeated 50 times, and the mean time (for all 1500 birds), standard deviation, minimum and maximum values were computed, as shown in Table 5.12. The mean was also divided by 1500 to calculate an indicative time for a single

Table 5.12: Classification times for 1,500 birds in seconds and the estimated classification times for a single bird in milliseconds.

	Appearance				Motion				Marini			
	NB	SVM	RT	RF	NB	SVM	RT	RF	NB	SVM	RT	RF
Mean/1500 Birds	13.43	13.43	12.48	14.10	18.01	18.07	16.28	19.94	7.67	7.72	6.98	7.91
Min/1500 Birds	12.05	12.05	11.98	13.01	16.56	16.61	14.68	17.42	5.91	5.96	4.72	6.61
Max/1500 Birds	17.37	17.36	16.38	18.42	21.16	21.25	19.42	22.98	14.05	14.16	12.98	15.98
σ/1500 Birds	0.96	0.96	0.84	1.01	1.33	1.34	1.14	1.49	1.73	1.75	1.69	1.98
Mean/Birds (in ms)	8.95	8.95	8.32	9.40	12.01	12.05	10.85	13.29	5.11	5.15	4.65	5.27

bird which represents "in the field" performance for the system.

From Table 5.11, it can be seen that using the appearance features, training took on average 157.51, 152.1, 99.16 and 114.21 seconds with the Naive Bayes, SVM, Random Tree and Random Forest respectively. Marini et al. (2013) was faster with 85.74, 87.45, 65.33 and 97.34 seconds. The motion features took 202.65, 202.72, 110.12 and 180.13 seconds respectively. As expected, the training times were slower with larger feature sets but very acceptable for off-line training. Comparison of the four classifiers (NB, RF, RT and SVM) shows little variation using the same features, and this illustrates clearly that it is the feature extraction process, rather than the training process, which dominates the time required for the training phase. The training complexity of SVM is of the order $\mathcal{O}(nd)$ (Keerthi et al., 2006), and Naive Bayes training complexity has been shown similarly to be of order $\mathcal{O}(nd)$ (Zheng and Webb, 2005), where n is the number of training samples and d is the feature dimension. The computational complexity of a random tree based on unpruned CART trees is of the order $\mathcal{O}(ndLog(n))$ (Hall et al., 2011). Random forest models are unpruned CART trees. Thus, if you grow M trees, then its computational complexity becomes $\mathcal{O}(M(ndLog(n)))$ $\mathcal{O}(M(mnLog(n))$ (Biau, 2012). Given the large set of features, we expect the complexity of the training phase to be approximately linear in the number of training samples, since both the dominant feature extraction process and training, are both linear.

From Table 5.12, the observation was that, Marini et al.'s returns a classification slightly faster than either appearance-based or motion-based classifiers. The classification complexity of SVM, Random Trees, Random Forest and Naive Bayes are all of the order $\mathcal{O}(d)$ (Zheng and Webb, 2005; Keerthi et al., 2006), which means that the time it takes

to classify a bird is linearly dependent on the feature dimension. However, the results indicate that the choice of a classifier makes relatively little difference to performance, which again is heavily dominated by the feature extraction process. Crucially, and despite an approximate doubling of processing time when compared with Marini et al.'s, both the motion and appearance-based feature sets are able to return a single classification in less than 10ms. Even when appearance and motion features are combined, a single bird is classified in approximately 20ms, which is more than suitable for real-time application.

5.5 Conclusion

This chapter described the use of the proposed motion and combined feature sets for automated classification of flying bird species. Four classifiers (Random trees, Random forest, Naive Bayes and SVM) were used to evaluate the proposed feature sets experimentally. Supporting experimental results in which these feature sets were utilised to classify species based on the thirteen classes (eleven bird species, one with three colour forms) datasets were also presented.

Experiments show that the RF classifier had the best classification rate across all datasets. Results from this classifier were used to perform various evaluations presented in this chapter. The automated classification of species using the motion features alone showed a more than 10% correct classification rate for most species. Thus establishing their usefulness for classification of species and hypothesising that they will be most effective for species at a distance, where coloured features are more likely to attenuate.

Observation of whether motion cues alone can help differentiate between species with similar appearance, revealed that they may not be sufficient to differentiate these species but may be used as a weak classifier. Even though motion cues can be used as a weak classifier, its classification rates suggests that it can help differentiate species with similar appearances. When appearance features alone are used to classify these species, more misclassifications are recorded compared with the combined (appearance and motion) features.

The experiments also showed that the three Budgerigar species, with nearly the same

flight pattern, were misclassified as the other Budgerigars. Precisely, more than half Budgerigar (yellow) and Budgerigar (wild-type) were misclassified as other Budgerigars and more than a third of Budgerigar (blue) were also misclassified as the other Budgerigars. This is probably because Budgerigars varies by colour and not motion and hence their low correct classification rates when only motion features are used.

The RF classifier with the thirteen classes dataset shows 83% more correct classification when compared with the minority class, which is only 2%. This may be due to the imbalanced nature of the dataset or it being redundant. Another observation of the results shows that species with small samples were also generally misclassified as species with larger samples (Black-headed Gulls and Common House Martin). In particular, Budgerigars had smaller samples and thus were in some cases misclassified as either Black-headed Gulls and Common House Martin.

The naive addition of the motion features to the appearance made little difference to the overall classifier performance when compared with only appearance. The combined features set increased the performance of the NB and SVM classifiers by 15% and 6% respectively, but made no difference to RF and decreased the performance of RT classifier by 3%. When experiments were performed with only motion features, the results show that they can be used as a weak classifier, thus, motivating the exploration of effective features selection in the following chapter.

Finally, the computational complexity of the models were compared with the state-of-the-art, through both empirical evaluations to determine if the algorithms are capable of running real-time. It was established that work in the state-of-the-art method by Marini et al. was slightly faster due to the smaller feature set used. The proposed algorithm was also shown to be capable of running comfortably in real-time.

The following chapter will explore the selection of important features using correlation- and classifier-based feature selection techniques. This is believe to help eliminate irrelevant and superfluous features and improve correct classification rates.

Chapter 6

Feature Selection for the Bird Species Problem

In the previous chapter, a comparison between experiments performed using the combined features and appearance was made. This revealed an increase in overall correct classification rate for the NB and SVM classifier and a small or no reduction in correct classification rates for the RT and RF classifier when the combined features were used. In particular, there was no increase in classification rate when RF classifier was used. The NB and SVM classifiers showed an increase in correct classification rate by 15% and 6% respectively whiles the RT classifier decreased it classification rate by 3%. The last experiment performed using only the motion features showed that they can be used as weak classifiers, thus motivating the exploration of feature selection to optimise classification performance. All the methods above use large feature sets (320 features), but the presence of redundant or irrelevant features may affect classification rates negatively (Hall, 1999; Yu and Liu, 2003). Thus we consider that feature (dimensionality) reduction and/or selection is an important aspect of the fine-grained bird species identification problem.

The aim of the work presented in this chapter is to eliminate redundant and irrelevant features from the combined feature set and to determine if correct classification rates can further be improved. Specifically, we explore and compare feature selection methods to remove redundant features, and show that using a refined subset of the original 320 features may significantly improve recognition rates on the dataset of thirteen classes.

The selected optimal features were used with the four standard classifiers to classify species and comparatively evaluate relative performance. We finally identify the most important motion features and determine the contribution made by these features to effectively classify species. The major contribution of this chapter is the use of feature selection and reduction techniques to improve correct classification rates of the combined feature set, and to demonstrate the relative importance of motion features within this reduced set. The chapter is organised into the following sections:

- Section 6.1 succinctly introduces the datasets used in this chapter and gives a brief overview of processing architecture used.

- Feature selection techniques are described in Section 6.2 and the experimental setup in Section 6.3.

- In Sections 6.4 and 6.5, evaluation and experimental results are presented, which includes a comparison of different classifiers and feature selection techniques, overall system performance, and concludes with an analysis of the relative contribution of motion features to effective recognition.

- Finally, conclusions are drawn to the chapter in Section 6.6, which include a brief introduction to what will be described in the next chapter.

6.1 Datasets, Methods and Preprocessing

The dataset used for experiments in this chapter is an extended set of videos covering thirteen classes made up of eleven unique species with one (Budgerigar (*Melopsittacus Undulatus*)), having three colour forms. This has a total of 957 videos, representing 103,393 image frames. This dataset has been described in detail in Chapter 4. This dataset was chosen for experiments in this chapter as it is an extensive dataset, representing more species than the other datasets.

For each video, birds' silhouettes were extracted using the background Gaussian mixture model proposed by Zivkovic and van der Heijden (2006) which was introduced

in Chapter 3. To detect the connected components, contours were obtained from the binary image using the contour algorithm proposed by Suzuki et al. (1985), also described in Chapter 3. An oriented bounding box was fitted to each silhouette and a selection of metrics (height, width and hypotenuse, centroid, silhouette and contour points) were measured. Colour moments, shape moments, greyscale histogram, Gabor filter and log-polar features were extracted from the segmented silhouettes to represent appearance features. The details of these have been presented in Chapter 4.

Before extracting motion features, the centroids of the segmented bird silhouettes were first extracted from each frame of the videos. The 2D centroid position was used to form a trajectory in the image frame, details of which have been presented in Chapter 5. Motion features, including curvature scale space (CSS), turn-based, wing-beat frequency, centroid distance function (CDF), vicinity and curvature based on sine and cosine were extracted from these trajectories.

Most of the appearance and motion features were represented as statistical features to provide information about the variability and the distribution of the data and also to enable classification in real time, as statistical features reduce the size of the feature set. The statistical features computed include the mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings. Again, details of these have been provided in Chapter 4. The experiments discribed in this chapter uses motion and/or appearance features which have also been detailed in Chapter 5.

6.2 Feature Selection Techniques

Two feature selection techniques were used to determine which features are most important in the combined feature set. One technique was selected from the filter methods and the other from the wrapper methods. This was because different methods return different features and comparing the results of different methods will ensure the right features are selected for the classification model. The filter methods can be used to select important

features ranked using classifier independent methods such as Fisher score, Pearson correlation coefficient, mutual information or relief. The wrapper methods use the machine learning algorithm to measure the worth of feature subsets. The filter method uses the correlation-based feature selection technique, which is a faster algorithm with the premise that good features are highly correlated with the class yet uncorrelated with each other. This method was selected as it is faster and classifier independent. The wrapper method used is the classifier-based feature selection method, which is based on the random forest classifier. This method was used for two reasons: firstly, the classifier used is the best performing classifier and secondly, it is a state-of-the-art wrapper technique. These two methods will be detailed in the following two sections.

6.2.1 Classifier-based Feature Selection Method

The classifier-based feature selection is based on the Random Forest classifier (Breiman (2001)). This method introduces a random permutation into the learning process, in order to produce multiple decision trees from a single dataset (thus forming a "forest"). Aggregation techniques, such as majority voting, are then used to combine the predictions from all of the trees. The method combines Breiman's "Bagging" (Breiman, 1996) whilst injecting random perturbations into the feature selection, for building a collection of decision trees with a controlled variance. The dataset is split into a training set and a test set, known as the *out of bag* cases, in a Random Forest classification problem.

The use of the Random Forest classifier for feature selection was first introduced by Breiman (2001). The algorithm have been briefly presented in Algorithms 8 and 9. There were two methods defined in his work, namely Gini importance and permutation importance. Gini importance is a fast variable importance method, which is based on the Gini impurity index (see Equation 6.1) used for the calculation of tree splits during training, and is obtained by adding up the Gini decrease for each feature over all trees in the forest.

$$G = \sum_{i=1}^{n_{classes}} p_i(1-p_i) \tag{6.1}$$

Where $n_{classes}$ is the number of classes in the target variable and p_i is the proportion of samples that belong to the class i. However, this thesis used permutation importance, which is widely used with Random Forests and it is also known to have better performance than Gini importance (Louppe et al., 2013). The variable importance of a feature using the permutation importance technique is computed as the mean decrease in the model's accuracy on the *out of bag* samples, when the values of the respective feature are randomly permuted.

In this case, taking $F = \{X_1, \ldots, X_N\}$ of numerical input features and Y categorical outputs; then, for some $X_i \in F$, we randomly permute the values of the variable X_i in the *out of bag* observations OOB, and put this down the tree. We then find the number of votes cast for the correct class for the permuted X_i data ($POOB_{X_i}$), and the number of votes for the correct class in the original *out of bag* data (OOB_{X_i}). We averaged the difference between the permuted and original over all the trees in the forest (n_{trees}), as given in equation (6.2), to the variable importance of the X_i variable M_{X_i}. All the variables were then arranged in descending order of variable importance (M_{X_i}) to perform the feature rankings presented in the experimental work.

$$M_{X_i} = \frac{1}{n_{trees}} \sum_{t=1}^{n_{trees}} (POOB^t_{X_i} - OOB^t_{X_i}) \tag{6.2}$$

Algorithm 8: Random Forest Algorithm. Adapted from Breiman (2001)

1 Draw n_{trees} bootstrap samples from the original data;
2 For each of the bootstrap samples, grow an unpruned classification tree, using bagging. Randomly sample the predictors and choose the best split from among those variables;
3 Predict new data by aggregating the predictions of the n_{trees} trees and using majority votes to determine class;

6.2.2 Correlation-based Feature Selection Method

The correlation-based method used is based on the work of Hall (1999), which was implemented in Hall et al. (2009). To use this method, a feature-to-class and feature-to-feature correlations matrices are first calculated from the dataset using Pearson's correlation $r_{f_1 f_2}$ with standardised variables, as described by Equation 6.3.

Algorithm 9: OOB Error Rate. Adapted from Breiman (2001)

1 **Algorithm** OOB Error Rate;
2 At each bootstrap iteration, predict "out-of-bag" (OOB) data using the tree grown with the bootstrap sample;
3 Aggregate the OOB predictions and calcuate the error rate (OOB estimate error rate);

$$r_{f_1 f_2} = \frac{\sum f_1 f_2}{n\sigma_{f_1}\sigma_{f_2}} \tag{6.3}$$

f_1 and f_2 are two continuous variables expressed in terms of deviations. The feature subset space is searched using a best-first search algorithm (see Algorithm 10 adapted from Hall (1999)) in order to determine the most effective subset of features. An advantage of this method is that if the path currently being explored begins to return less promising results, the best first search can backtrack to a previous subset and continue the search from there. The *merits* of each feature subset are then computed using equation 6.4.

$$M_s = \frac{k\bar{r_{cf}}}{\sqrt{k+k(k-1)\bar{r_{ff}}}} \tag{6.4}$$

M_s is the merit of feature subset containing k features, $\bar{r_{cf}}$ is the mean feature-class correlation ($f \in S$) and $\bar{r_{ff}}$ is the average feature-to-feature inter-correlation. In this context, the numerator gives an indication of how predictive groups of features are, and the denominator represents the level of redundancy among them. Now, assuming the subset $S = \{f_1, f_2, f_3, f_4 \ldots f_k\}$ contains k features, then the feature-to-feature correlation can expressed as Equation 6.5.

$$\bar{r_{ff}} = \frac{\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} r_{ff}\{f_i, f_j\}}{{}^{k}C_2} \tag{6.5}$$

$r_{ff\{f_i,f_j\}}$ is the pairwise correlation of feature f_i with f_j, and ${}^{k}C_2$ is the number of pairwise combination that is possible from the subset S without repetitions. Also, the feature-class correlation can be calculated using Equation 6.6

$$r_{cf}^{-} = \frac{\sum_{i=0}^{k} r_{cf}\{f_i, c_i\}}{k} \tag{6.6}$$

Where $r_{fc\{f_i,c_j\}}$ is the pairwise correlation of feature f_i with class c_j. All the variables were then arranged in descending order of merit (M_s) to perform the feature selection experiment.

Algorithm 10: Best First Search Algorithm. Adapted from Hall (1999)

```
1  OPEN ← startstate;
2  CLOSED ← ∅;
3  BEST ← startstate;
4  Lets = arg max e(x);
5  OPEN ← OPEN − s;
6  CLOSED ← CLOSED + s;
7  if e(x) ≥ e(x) then
8  |   BEST ← s
9  end
10 For each child t of s ∉ OPEN or CLOSED;
11 OPEN ← OPEN + t;
12 if BEST changed then
13 |   goto 4
14 else
15 |   Return BEST
16 end
```

Tables 6.1 and 6.2 have been used to exemplify these. In order to calculate the heuristic merits (M_s), the correlation matrix (Table 6.1) for feature-feature and feature-class are first calculated using the Pearson correlation Equation 6.3. Heuristic merits (M_s) are then calculated using the correlation matrix. Note that the feature-feature and feature-class correlations for at most two features can be referenced directly from Table 6.1. For example,

Table 6.1: Example correlation matrix calculated with sample features from dataset #2.

	a	b	c	d	e	f	**class**
a	1.000	0.096	0.112	0.033	0.043	0.053	**0.122**
b		1.000	0.142	0.022	0.156	0.177	**0.185**
c			1.000	0.311	0.145	0.233	**0.324**
d				1.000	0.033	0.133	**0.135**
e					1.000	0.000	**0.223**
f						1.000	**0.034**

the subset $\{a\}$, has $k = 1$, feature-feature corelation $r_{ff} = 1.000$ and feature-class correlation $r_{cf} = 0.122$ (see Table 6.1), using Equation 6.4, the heuristic merits $M_s = 0.122$. However, when the subset is greater than two then the feature-feature correlation can be derived by first permuting the feature subset and referencing their correlation from the matrix and taking the mean of the results. The search starts with the empty set which has a merit of zero. The best subset is then choosen based on the computed merits. These values have been bolded in the Table 6.2. Assuming the subset $S = \{f_1, f_2, f_3, f_4 \ldots f_k\}$ contains k features, then the feature-feature correlation can be expressed as Equation 6.5. For example, the subset $\{b, c, e\}$ has $k = 3$, feature-feature corelation $(r_{ff}) = 0.443$ which is calculated using Equation 6.5 and feature-class correlation $(r_{cf}) = 0.732$ also calculated using Equation 6.6. Finally the heuristic merits $M_s = 0.923$ can be computed using Equation 6.4.

According to Hall (2000) this method has a low time complexity. There are $m((n^2 - n)/2)$ operations required to calculate the pairwise feature correlation matrix. Where m is the number of instances and n is the initial number of features. In a worse case scenario, the feature selection search requires $(n^2 - n)/2$ operations for the forward selection. Also, k additions are required in the numerator and $(k^2 - k)/2$ addition in the denominator for evaluating heuristic merits containing a subset S with k features. Finally, it has been shown by Hall (2000) that this method is faster and comparable in classification performance with the various wrapper methods.

6.3 Experimental Setup

Five sets of experimental evaluations are detailed in this section. They include:

Table 6.2: A Best First search using the correlations in 6.1.

	k	$\bar{r_{cf}}$	$\bar{r_{ff}}$	M_s
{}	0	-	-	0
{a}	1	0.122	1.000	0.122
{b}	1	0.185	1.000	0.185
{c}	**1**	**0.324**	**1.000**	**0.324**
{d}	1	0.135	1.000	0.135
{e}	1	0.223	1.000	0.223
{f}	1	0.034	1.000	0.034
{a,c}	2	0.446	0.112	0.598
{b,c}	2	0.509	0.142	0.674
{c,d}	2	0.459	0.311	0.567
{c,e}	**2**	**0.547**	**0.145**	**0.723**
{c,f}	2	0.358	0.233	0.456
{a,c,e}	3	0.669	0.3	0.916
{b,c,e}	**3**	**0.732**	**0.443**	**0.923**
{c,d,e}	3	0.682	0.489	0.840
{c,e,f}	3	0.581	0.511	0.708

- The presentation of results that identify the most important feature subsets after removal of irrelevant or redundant features using the two feature selection techniques.

- Revisiting the appearance feature results presented in Chapter 4 but using the 68 optimal features selected by the classifier-based method in Section 6.4.

- Revisiting the motion feature results presented in Chapter 5 but using the 12 optimal features selected by the classifier-based method in Section 6.4.

- Revisiting the combined feature results presented in Chapter 5 but using the 80 optimal features selected by the classifier-based method in Section 6.4.

- Finally, the most important motion features were identified, and their contribution to the overall classifier effectiveness quantified.

All experiments were performed on a Mac book pro laptop running OS X 10.9.5, with 2.5 GHz Processor and 4 GB RAM. The pre-processing and feature extraction algorithms were all implemented in C++ with XCode 5.1.1 and OpenCV 3.0, whilst the classification and feature selection algorithms were implemented in WEKA 3.7 (Hall et al., 2009).

A five-fold cross-validation scheme was used for all experiments performed, in which the complete dataset was randomly partitioned into five equal-sized subsets. Five iterations were then performed: for each iteration a unique subset was retained as a validation set and the others used for training. The mean of the results from each fold was taken as the overall result.

Each experimental run sampled individual image frames (from the training and test set) from which the corresponding appearance and motion features were extracted. These features were concatenated to form the full combined feature set, comprising of 320 features (169 appearance and 151 motion features). All features were stored in a WEKA compatible format.

Finally, the feature set were loaded into WEKA for classification and feature selection. Four classifiers: Naive Bayes (NB), Random Forest (RF), Random Tree (RT) and Support Vector Machine (SVM) were used to perform the classification experiments, and the results reported for each.

The Naive Bayes classifier assumes a Gaussian mixture model over the whole training data distribution, with one component per class, and parameters were estimated from the training data.

The SVM classifier was based on LibSVM proposed by Chang and Lin (2011), which is comparable to that used by Marini et al. (2013), and implemented using a radial basis function kernel, with the gamma and cost parameters optimised using a 5-fold grid search. In the case of the Random Tree classifier, K randomly chosen attributes at each node were considered , in this case $K = int(log_2(\#features) + 1)$, and the maximum depth of the tree was set to be *unlimited*. Twenty trees were used for the Random Forest classifier, as this results in a convergence of the *out of bag* errors (other parameters for this classifier are the same as for the RT).

For feature selection, the merits M_s (see Equation 6.4) for each of the 320 features were sorted in descending order. Starting with the complete set of 320 features, the 10 least significant features were iteratively removed, until only 10 remained; for each iteration the mean classification rate obtained with the corresponding feature subset was used

to plot a learning curve. This was repeated for each classifier, and helped to identify the optimal number of features in each case.

The merits M_i (see Equation 6.2) for each of the 320 features were sorted in descending order, and again, starting with the full set of 320 features, the 10 least significant features were iteratively removed and the remaining set were used to plot a learning curve. Again, this was repeated for each classifier which enabled the identification of the optimal number of features.

6.4 Optimal Feature Selection

The hypothesis is that the use of combined appearance and motion features may be used to improve classification performance; however, the results in the previous chapter demonstrate that simply combining the appearance and motion features does not automatically improve the correct classification rates. In Chapter 5 it was hypothesised that performance of the combined set may be undermined by the presence of redundant features and that significantly better classification rates may be achieved using a subset of the full 320 features. The results of the experimental investigation of optimal feature selection are presented in this section.

As previously mentioned, two methods of feature selection were evaluated: the correlation-based method proposed by Hall (1999), and a variation of the machine learning based method proposed by Breiman (2001). Using each method, the optimal feature subset were determined, and the performance of each of the four classifiers using that subset was evaluated. This section describes how the optimal subset is determined, and also describes the composition of that subset (ie which of the original features are retained). The following procedure was used to determine the optimal feature subset, for both correlation and classifier-based methods:

- The feature selection method was used to rank features, in order from the most to least effective
- For consistency, the 10 lowest ranked features were iteratively removed from the list

- For each iteration, the performance of each of the four classifiers were evaluated using the remaining features
- This was repeated until only the 10 highest ranking features remained
- A graph of the correct classification rates against the number of features were plotted, and the maxima for each of the classifiers were estimated.

Figures 6.1 and 6.2 show the resulting graphs for the correlation and classification based methods respectively. Each classifier is shown as a separate curve, and the mean classification rate across all classifiers is also shown.

Peaks in the classifier-based method occur at 180 features for the NB classifier, 320 for SVM, and 70 for RT and RF respectively. Likewise, peaks using the correlation-based method occurred at 170 features for the NB classifier, 320 for SVM, and 70 for RT and RF classifiers respectively.

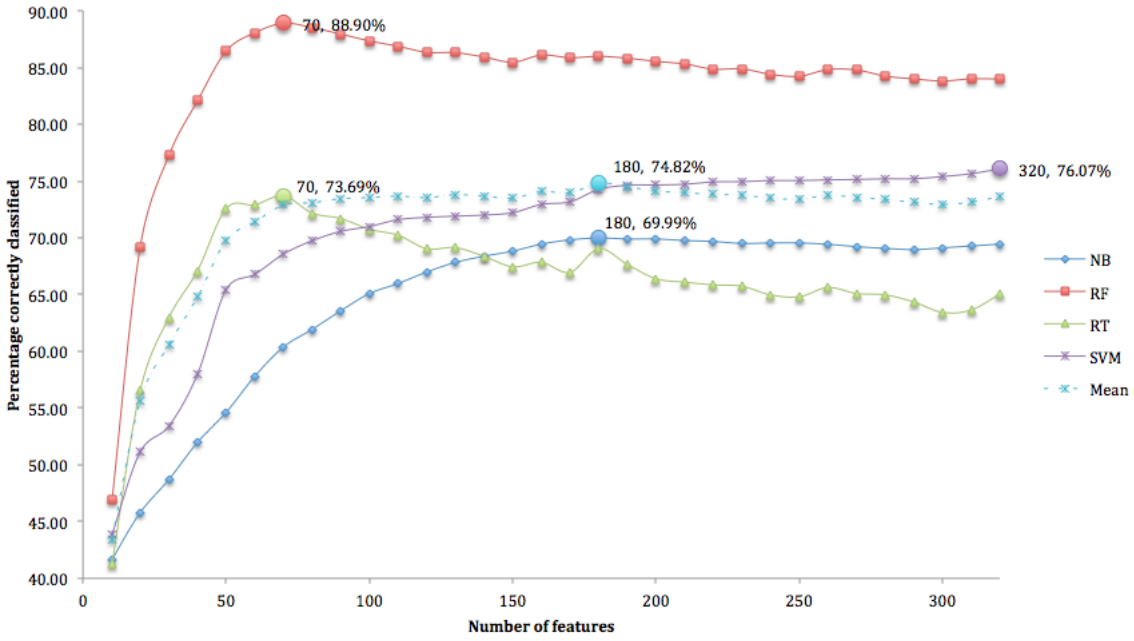


Figure 6.1: Plot of correct classification rates vs. number of features for the four standard classifiers when classifier-based selection is applied. The maximum for each classifier is marked with a solid circle, and labelled with the number of features and correct classification rate.

The mean of the four curves were taken and this was plotted with dashed lines: the highest correct classification rate for the mean curve is 75.40% when the correlation-based

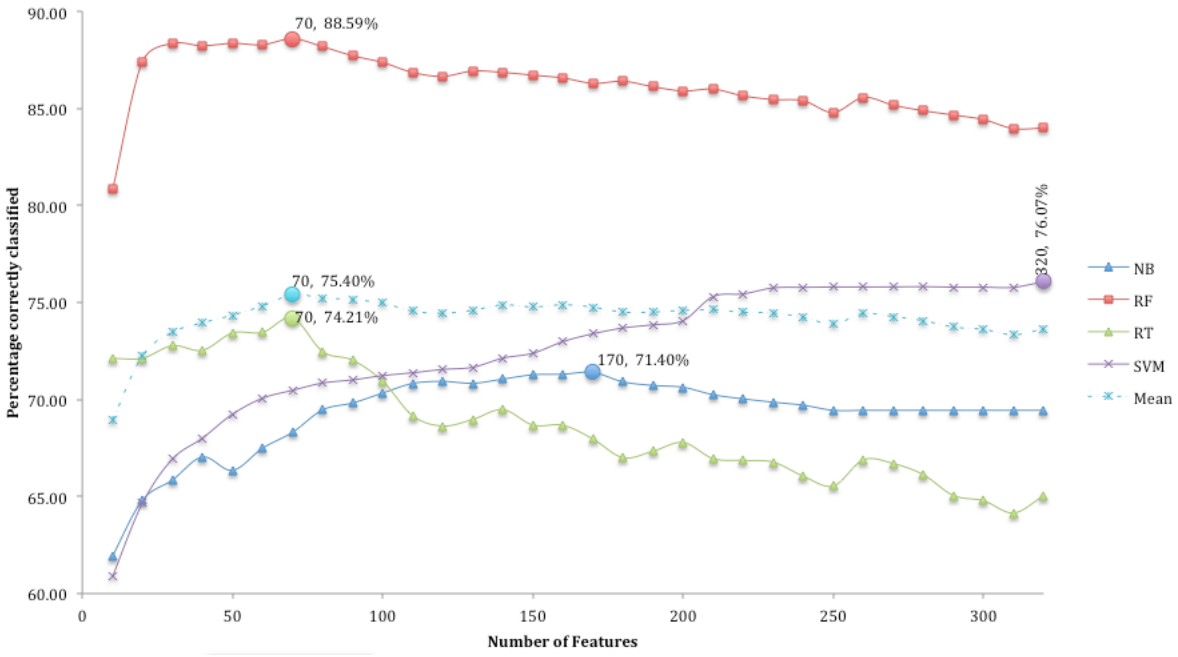


Figure 6.2: Plot of correct classification rates vs. number of features for the four standard classifiers when correlation-based selection is applied. The maximum for each classifier is marked with a solid circle, and labelled with the number of features and correct classification rate.

method is applied (this corresponds to 70 features). Similarly, this occurs at 74.82% when the classifier-based method is applied, which is at 180 features. The mode of both feature selection techniques occurred at 70 features (In both techniques RT and RF classifiers peaked at 70 features). Therefore, based on the mode and mean, the most optimal correct classification rates are achieved using the subset of 70 highest-ranked features (for both feature selection methods).

Table 6.3 shows feature groups by type, before and after selection, for both correlation-based (CoBfs) and classifier-based (CBfs) methods. From the table, the classifier-based method selected 62 appearance features, from seven feature groups, and 8 motion features from two groups. The correlation-based method selected 49 appearance features from six groups, and 21 motion features from three groups. It was noted that for the motion features, wing beat frequency and vicinity features were selected irrespective of the method used. This suggests that wing beat features can effectively contribution to species classification since they were selected by both methods. The features from the correlation-based method were used for the remaining experiments since the highest

Table 6.3: The number of features remaining in each feature group before and after applying the classifier-based (CBfs) and correlation-based (CoBfs) Feature Selection (FS) methods. The table also includes the top feature selected for each of the classes of features, using both feature selection methods.

	Feature Group	# before FS	# after CBfs	# after CoBfs	**Top Selected Feature (CBfs)**	**Top Selected Feature (CoBfs)**
Appearance	Hue color features	37	13	18	σ of Hue	Mean of Hue
	Saturation colour features	35	16	12	Mean of Saturate	Mean of Saturate
	Value colour features	37	28	16	Entropy of value	Entropy of value
	Shape	17	1	1	Hu's First invariant	Hu's First invariant
	Gabor	20	1	1	Mean of Gabor (at $\theta = 0$)	Mean of Gabor (at $\theta = 0$)
	Grayscale	8	1	0	Mean of Grayscale	N/A
	LogPolar	15	2	1	Mean of Logpolar hue	Mean of Logpolar hue
Motion	FFT (Wingbeat)	27	7	8	First Peak of FFT (width)	First Peak of FFT (width)
	CSS	22	0	0	N/A	N/A
	CDF	10	0	0	N/A	N/A
	Turn	62	0	12	N/A	Turn ($\theta_{i=55}$)
	Vicinity	20	1	1	Mean of Vicinity Curliness	Mean of Vicinity Curliness
	Curvature	10	0	0	N/A	N/A
	Total Features	**320**	**70**	**70**		

mean correct classification rate (75.40%) was achieved with this method.

The hue colour feature's top selected feature using the classifier- and correlation-based methods is the hue histogram's standard deviation and mean respectively. The mean of the hue histogram describes the general brightness of the hue colour, whereas the standard deviation (σ) describes the contrast. The top selected saturation feature for both feature selection methods is the saturation histogram's mean, which also describes the general brightness of the saturation colour. Finally, for the colour features, the top selected value feature for both the classifier- and correlation-based method is the value histogram's entropy, which shows how many bits are needed to code the image data.

The most important feature for identifying the shape of bird species is Hu's first moment as depicted by both correlation- and classifier-based feature selection methods. The Hu's first moment determines the shape of the bird irrespective of translation, rotation, or scale.

Finally, most important motion feature is wingbeat frequency, which is represented by FFT. The top selected feature for this class of feature is the first FFT peak computed using the width metric. This is the wingbeat frequency of the bird species.

6.5 Discussion of Results

This section presents more detailed results obtained using the 70 selected features by the correlation-based method in Section 6.4, as this is overall the best performing subset. Tables 6.4, 6.6 and 6.9 summarise the correct classification rate for each species, using the four standard classifiers with the appearance, motion and combined features respectively: the result for *combined* features are obtained from the complete subset of 70 features (mixed appearance and motion), the results shown for *appearance* are obtained using only the 49 appearance features from that subset; whereas the result for *motion* features are obtained using only the 21 motion features from that subset. A confusion matrix was also provided for the RF classifier (the best performing classifier) in Table 6.5, 6.7, and 6.10 for the appearance, motion and combined feature sets respectively, in order to help explore misclassification among species with similar appearances and those with closely related motion. The cross-species confusion matrix for the other classifiers have been presented in Appendix D, E and F for the appearance, motion and combined feature sets respectively.

6.5.1 Appearance Feature Set Results Revisited

This section revisits results presented in Chapter 4 using the 49 appearance features selected by the classifier-based method in Section 6.4. The correct classification rates of species with fine-grained appearances (closely related species) and all species within the dataset of thirteen classes are examined.

Again, it is evident from Table 6.4 that the performance of the RF and RT classifiers remain superior to both NB and SVM. The Random Forest classifier gives the highest correct classification rate (85%) with this dataset. Particularly, the RF classifier outperforms NB by between 7-54% across all species. Similarly, the Random Tree classifier also had a better performance than either SVM or NB classifier.

To confirm that the Random Forest (RF) classifier's correct classification rate is statistically significant compared with the other classifiers, the Wilcoxon's test statistic W is calculated. For RF compared with NB classifiers, $W = 0$ (the smaller of $W^+ = 645$ and

Table 6.4: Summary of species correct classification rates based on **the appearance features** after the correlation-based feature selection was performed and using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in Chapter 4.

Species	NB	RF	RT	SVM
a=Alexandrine Parakeet	29% ± 0.81%	83% ± 0.75%	66% ± 0.87%	53% ± 0.03%
b=Nanday Parakeet	46% ± 0.37%	82% ± 0.29%	62% ± 0.81%	39% ± 0.4%
c=Blue-crowned Parakeet	53% ± 0.04%	85% ± 0.43%	71% ± 0.45%	49% ± 0.43%
d=Common House Martin	86% ± 0.19%	93% ± 0.21%	88% ± 0.59%	96% ± 0.49%
e=Eastern Rosella	53% ± 0.62%	80% ± 0.68%	62% ± 0.87%	39% ± 0.38%
f=Budgerigar (yellow)	64% ± 0.51%	86% ± 0.61%	67% ± 0.42%	67% ± 0.37%
g=House Sparrow	42% ± 0.15%	72% ± 0.51%	56% ± 0.25%	35% ± 0.25%
h=Budgerigar (wild-type)	19% ± 0.82%	62% ± 0.72%	46% ± 0.22%	14% ± 0.11%
i=Common Wood Pigeon	46% ± 0.13%	72% ± 0.08%	60% ± 0.05%	32% ± 0.68%
j=Black-headed Gull	70% ± 0.78%	96% ± 0.48%	91% ± 0.81%	94% ± 0.18%
k=Cockatiel	45% ± 0.52%	84% ± 0.72%	67% ± 0.54%	52% ± 0.89%
l=Budgerigar (blue)	38% ± 0.69%	79% ± 0.09%	62% ± 0.29%	39% ± 0.74%
m=Common Starling	54% ± 0.39%	76% ± 0.31%	58% ± 0.57%	72% ± 0.64%
Overall Correctly Classified	**57% ± 0.32%**	**85% ± 0.42%**	**73% ± 0.54%**	**65% ± 0.38%**

$W^- = 0$), when RF and RT are compared, $W = 0$ (this is the smaller of $W^+ = 553$ and $W^- = 0$) and $W = 0$ (the smaller of $W^+ = 672.5$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF classifier resulted in a W less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB classifier. This is also true for the RF compared with the RT and SVM classifiers, as both resulted in a W which is less than 415. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the seven species dataset is used with the appearance feature sets.

Majority and Minority Sampled Species

Classification rates across all species using the selected feature set improved for some of the species. Particularly, the maximum classification rate using the appearance features is 96% and the minimum is 62%, which is about 3% lower and 17% higher than when all feature sets were used respectively. The Black-headed Gull class had the highest classification rate (96%), corresponding to the majority class whilst the Budgerigar (wild-type) (62%) class had the lowest.

Table 6.5: The confusion matrix based on the **Random Forest** classifier with the **selected appearance features (49 features)**. The results are based on the correlation-based technique.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	82.8%	5.5%	1.9%	0.3%	0.3%	1.4%	0.8%	1.4%	0.4%	1.2%	1.2%	2.6%	0.3%	83%	12801
b=Nanday Parakeet	7.3%	82.3%	2.0%	0.7%	0.5%	1.6%	0.2%	0.9%	1.9%	0.5%	0.6%	0.7%	0.8%	82%	10025
c=Blue-crowned Parakeet	5.9%	2.1%	85.0%	0.4%	1.0%	1.1%	0.7%	1.2%	0.3%	0.3%	0.2%	1.4%	0.4%	85%	9076
d=Common House Martin	0.3%	0.2%	0.2%	93.4%	0.0%	0.0%	2.4%	0.2%	0.1%	0.7%	0.4%	0.0%	2.0%	93%	25517
e=Eastern Rosella	1.2%	1.5%	1.5%	0.4%	80.5%	0.8%	1.3%	1.0%	0.2%	4.0%	1.8%	5.2%	0.5%	80%	5929
f=Budgerigar (yellow)	3.6%	2.4%	0.8%	0.4%	0.3%	85.7%	1.4%	1.6%	0.2%	1.9%	0.6%	0.8%	0.3%	86%	7667
g=House Sparrow	0.4%	0.1%	0.1%	10.2%	0.3%	0.7%	72.0%	0.5%	0.1%	4.1%	1.3%	1.7%	8.5%	72%	10191
h=Budgerigar (wild-type)	7.4%	4.1%	3.2%	1.9%	1.2%	5.4%	3.3%	62.3%	1.3%	2.9%	2.2%	3.3%	1.5%	62%	6283
i=Common Wood Pigeon	2.3%	7.1%	2.6%	2.9%	0.7%	1.0%	0.7%	1.1%	72.1%	1.5%	2.3%	1.1%	4.6%	72%	4301
j=Black-headed Gull	0.3%	0.0%	0.0%	1.1%	0.1%	0.0%	1.0%	0.1%	0.0%	95.5%	0.2%	1.7%	0.0%	96%	38764
k=Cockatiel	0.7%	0.3%	0.1%	3.9%	0.4%	0.4%	1.4%	0.4%	0.1%	2.4%	84.2%	1.0%	4.8%	84%	9398
l=Budgerigar (blue)	2.4%	0.6%	0.4%	0.8%	1.6%	0.7%	1.4%	1.1%	0.0%	10.2%	1.4%	78.7%	0.7%	79%	12090
m=Common Starling	0.4%	0.5%	0.2%	7.1%	0.1%	0.6%	10.2%	0.2%	0.2%	0.8%	2.9%	0.4%	76.5%	76%	9865
Overall Correctly Classified														**85%**	

Appearance-Related Species

Given that the RF classifier has shown best overall performance, a detailed cross-species confusion matrix for this classifier was obtained using the appearance features with the larger dataset which was presented in Table 6.5. As already noted, Alexandrine Parakeet, Nanday Parakeet and Blue Crowned Conure have very similar appearances. It therefore appears that Alexandrine Parakeets are typically misclassified as Nanday Parakeets, Blue Crowned Conures as Alexandrine Parakeets and Nanday Parakeets as Alexandrine Parakeets when relying on appearances only. Surprisingly, there was decrease in the misclassification of these species when the best appearance features were utilised. For example, 1.4% less Alexandrine Parakeets were misclassified as Nanday Parakeets, 2.2% less Blue Crowned Conures were misclassified as Alexandrine Parakeets and the nisclassification of Nanday Parakeets as Alexandrine Parakeets increase by just 0.1%.

6.5.2 Motion Feature Set Results Revisited

This section revisits results presented in Chapter 5 using the 21 selected motion features by the correlation-based method in Section 6.4. The results were evaluated with those presented in Chapter 5, to ascertain if the overall correct classification rate increases across all classifiers and that the misclassification among species with fine-grained appearances (closely related species) improves. The results were also used to evaluate misclasification among the species with closely related motion (the three Budgerigar forms).

Using the selected motion features for classification, the performance of the RF and NB classifiers are superior to both RT and SVM (see Table 6.6), compared to the full

Table 6.6: Summary of species correct classification rates based on **the motion features** after the correlation-based feature selection was performed and using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in Chapter 4.

Species	**NB**	**RF**	**RT**	**SVM**
a=Alexandrine Parakeet	20% ± 0.17%	26% ± 0.82%	16% ± 0.01%	13% ± 0.17%
b=Nanday Parakeet	24% ± 0.61%	19% ± 0.02%	14% ± 0.18%	16% ± 0.42%
c=Blue-crowned Parakeet	17% ± 0.08%	13% ± 0.24%	11% ± 0.09%	17% ± 0.24%
d=Common House Martin	37% ± 0.15%	55% ± 0.30%	35% ± 0.17%	25% ± 0.35%
e=Eastern Rosella	6% ± 0.49%	16% ± 0.51%	8% ± 0.35%	15% ± 0.84%
f=Budgerigar (yellow)	13% ± 0.52%	20% ± 0.71%	12% ± 0.61%	16% ± 0.48%
g=House Sparrow	45% ± 0.22%	26% ± 0.51%	18% ± 0.21%	12% ± 0.62%
h=Budgerigar (wild-type)	9% ± 0.32%	18% ± 0.81%	11% ± 0.41%	15% ± 0.53%
i=Common Wood Pigeon	11% ± 0.62%	21% ± 0.49%	7% ± 0.89%	20% ± 0.66%
j=Black-headed Gull	69% ± 0.51%	82% ± 0.23%	58% ± 0.74%	69% ± 0.29%
k=Cockatiel	21% ± 0.11%	19% ± 0.27%	12% ± 0.29%	21% ± 0.36%
l=Budgerigar (blue)	32% ± 0.08%	17% ± 0.16%	19% ± 0.85%	14% ± 0.14%
m=Common Starling	53% ± 0.23%	30% ± 0.67%	20% ± 0.84%	28% ± 0.31%
Overall Correctly Classified	**40% ± 0.46%**	**44% ± 0.65%**	**30% ± 0.42%**	**33% ± 0.47%**

motion features (see Table 5.2), which had the RF and SVM classifiers outperforming both RT and NB by a small margin. The RF classifier gives the highest correct classification rate. Particularly, outperforming the other classifiers by between 4-14% of correct classification rate.

In Chapter 5, using motion cues alone resulted in a significant decrease in the correct classification rate of all classifiers, when compared with appearance features. The performance of the motion features with the RF classifier was between 13% - 18% across all species. In this section, the results of the selected motion features show an increase in correct classification rates for all four classifiers when compared with the full set of motion features. Specifically, there was an increase of 8%, 6%, 5% and 1% respectively when the NB, RF, RT and SVM classifiers are used but these results are still lower when compared with appearance features. Comparing the selected motion feature results with selected appearance set, there was a decrease in correct classification rates with all classifiers by between 17-43%.

Again, the RF classifier has shown best overall performance, therefore a detailed cross-species confusion matrix for this classifier obtained using the motion features is presented in Table 6.7. This result together with those in Chapter 5 (Table 5.3) were used

Table 6.7: The confusion matrix based on the **Random Forest** classifier with the **selected motion features (21 features)**. The results are based on the correlation-based technique.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	26.2%	9.0%	6.1%	21.3%	1.2%	1.2%	2.7%	0.6%	0.6%	20.8%	3.7%	2.0%	4.6%	26%	7845
b=Nanday Parakeet	19.1%	19.4%	6.8%	20.0%	0.9%	1.2%	2.9%	0.2%	0.3%	17.7%	4.3%	1.8%	5.4%	19%	6246
c=Blue-crowned Parakeet	19.3%	13.0%	13.2%	20.4%	1.0%	0.8%	3.2%	0.5%	0.3%	15.2%	4.7%	2.0%	6.4%	13%	5332
d=Common House Martin	4.6%	2.9%	1.6%	54.9%	0.3%	0.8%	1.2%	0.3%	0.1%	28.8%	1.1%	2.2%	1.3%	55%	17896
e=Eastern Rosella	14.4%	8.8%	5.7%	20.3%	16.5%	1.0%	2.2%	0.4%	0.3%	20.9%	3.2%	2.3%	4.1%	16%	3247
f=Budgerigar (yellow)	6.1%	4.6%	2.7%	23.3%	0.7%	20.2%	2.8%	1.4%	0.3%	26.9%	1.7%	7.0%	2.3%	20%	4329
g=House Sparrow	8.4%	5.1%	3.0%	22.7%	0.5%	1.3%	25.7%	0.8%	0.2%	23.5%	1.3%	3.0%	4.4%	26%	5318
h=Budgerigar (wild-type)	8.4%	5.6%	3.0%	19.6%	0.6%	3.0%	5.6%	17.6%	0.1%	24.6%	1.9%	6.8%	3.1%	18%	3349
i=Common Wood Pigeon	10.5%	7.1%	7.0%	18.7%	0.8%	1.3%	3.5%	0.9%	21.0%	19.2%	3.1%	2.3%	4.6%	21%	2027
j=Black-headed Gull	2.1%	1.2%	0.8%	10.9%	0.2%	0.4%	0.5%	0.2%	0.0%	81.8%	0.8%	0.8%	0.3%	82%	29695
k=Cockatiel	15.0%	9.2%	5.8%	18.0%	0.7%	1.1%	2.4%	0.6%	0.5%	22.4%	19.0%	1.8%	3.5%	19%	5687
l=Budgerigar (blue)	6.3%	3.5%	1.9%	29.4%	0.6%	4.2%	2.7%	1.1%	0.1%	30.1%	1.1%	17.0%	1.9%	17%	7030
m=Common Starling	12.8%	8.7%	5.6%	20.8%	0.7%	0.9%	4.3%	0.5%	0.3%	10.8%	2.9%	2.0%	29.7%	30%	5392
Overall Correctly Classified														**44%**	

to explore misclassifications in species with similar appearances.

Appearance-Related Species

When the selected motion features alone were used to classify species, it appears that misclassifications among species with similar appearances were reduced compared with motion features without feature selection. For example, 0.7% less Alexandrine Parakeets were misclassified as Nanday Parakeets; 0.8% less Blue-crowned Parakeets as Alexandrine Parakeets when the selected motion features were used. The misclassification of Nanday Parakeets as Alexandrine Parakeets remain unchanged. This suggests that motion cues may be used as a weak classifier to help differentiate these species and therefore this has been explored in the next section.

Majority and Minority Sampled Species

Using the selected optimal motion features, the RF classifier showed 82% correct classification rate for the majority class (class with most samples) with the thirteen classes dataset. This is 61% more compared with the correct classification rate of the minority class. The difference in correct classification rate between the majority and minority classes, in this case, is smaller compared with when the full motion features are used. This further confirms that the dataset is imbalanced.

Motion-Related Species

Misclassification of the three Budgerigar using motion features was not very obvious, as one may suspect an artifact of having much more samples for other species might be the

Table 6.8: The confusion matrix based on the Random Forest classifier with the **selected motion features (12 features)**. The results are based on the **correlation-based technique** and the **three Budgerigar forms**.

	f	**h**	**l**	**%CC**	**Samples**
f=Budgerigar (yellow)	43.1%	7.5%	49.4%	43%	4329
h=Budgerigar (wild-type)	17.6%	37.0%	45.3%	37%	3349
l=Budgerigar (blue)	17.5%	6.9%	75.6%	76%	7030
Overall Correctly Classified				**57%**	

cause for this. The results revealed that the three Budgerigar species were mainly misclassified as other Budgerigar. Specifically, 1.4% and 7.0% Budgerigar (yellow) were misclassified as Budgerigar(wild-type) and Budgerigar (blue) respectively, 3.0% and 6.8% Budgerigar (wild-type) were misclassified as Budgerigar (yellow) and Budgerigar (blue) respectively, and 4.2% and 1.1% Budgerigar (blue) were misclassifies as Budgerigar (yellow) and Budgerigar (wild-type) respectively. Surprisingly, this reduced after feature selection, which means some misclassification among Budgerigars were also due to redundancy in the data.

To further confirm that Budgerigars are misclassified as other Budgerigars, we performed classification using only the species with different colour forms (Budgerigar) based on the RF classifier and motion features alone. The results have been presented in Table 6.8. The results show that there were many misclassifications among these species. Particularly, 49.4% of Budgerigar (yellow) were misclassified as Budgerigar (blue), 45.3% of Budgerigar (wild-type) were misclassified as Budgerigar (blue), 17.6% of Budgerigar (wild-type) were misclassified as Budgerigar (yellow) and 17.5% of Budgerigar (blue) were misclassified as Budgerigar (yellow). However, when all classes were used in the classification, there was more misclassification of these species as Black-headed Gulls and Common House Martin, which have much more samples than other species. The reason may therefore be attributed to having more Black-headed Gull data than all the other species.

Table 6.9: Summary of species correct classification rates based on **the combined features** after the correlation-based feature selection was performed and using the four standard classifiers. These results use the **thirteen classes (eleven bird species, one with three colour forms)** introduced in Chapter 4.

Species	NB	RF	RT	SVM
a=Alexandrine Parakeet	39% ± 0.65%	85% ± 0.63%	64% ± 0.45%	56% ± 0.08%
b=Nanday Parakeet	53% ± 0.17%	84% ± 0.35%	61% ± 0.43%	44% ± 0.19%
c=Blue-crowned Parakeet	58% ± 0.24%	87% ± 0.65%	69% ± 0.47%	52% ± 0.71%
d=Common House Martin	91% ± 0.87%	98% ± 0.27%	89% ± 0.86%	97% ± 0.35%
e=Eastern Rosella	57% ± 0.18%	81% ± 0.83%	59% ± 0.02%	34% ± 0.49%
f=Budgerigar (yellow)	72% ± 0.61%	86% ± 0.72%	63% ± 0.54%	73% ± 0.31%
g=House Sparrow	57% ± 0.82%	71% ± 0.86%	53% ± 0.44%	40% ± 0.65%
h=Budgerigar (wild-type)	31% ± 0.36%	62% ± 0.29%	44% ± 0.71%	26% ± 0.87%
i=Common Wood Pigeon	50% ± 0.87%	67% ± 0.08%	51% ± 0.19%	22% ± 0.79%
j=Black-headed Gull	82% ± 0.28%	99% ± 0.53%	92% ± 0.53%	96% ± 0.13%
k=Cockatiel	54% ± 0.29%	85% ± 0.42%	65% ± 0.27%	58% ± 0.71%
l=Budgerigar (blue)	57% ± 0.69%	79% ± 0.66%	63% ± 0.15%	41% ± 0.73%
m=Common Starling	66% ± 0.77%	76% ± 0.56%	57% ± 0.22%	72% ± 0.63%
Overall Correctly Classified	**68% ± 0.52%**	**89% ± 0.53%**	**74% ± 0.41%**	**71% ± 0.51%**

6.5.3 Combined Feature Set Results Revisited

This Section revisits the combined feature set results presented in Chapter 5 by re-classifying species using the features selected by the correlation-based method in Section 6.4. These results were evaluated to ascertain if the overall correct classification rate increases across all classifier and if misclassification among species with fine-grained appearances improve when the selected combined features are used.

From Tables 5.7 and 6.9, the change in classification rates are computed. The results show an increase of 5%, 9% and 6% with RF, RT and SVM classifiers respectively, and a 1% decrease with NB. This is significant, as the preferred classifier shows an overall correct classification rate of 89%. The other classifiers, RT, NB and SVM achieved rates of 74%, 60% and 60% respectively.

Appearance-Related Species

These results show a further decrease in the instances misclassified among species with closely related appearances, which may be attributed to the reduction of irrelevant and redundant features. For example, using the RF classifier, the misclassification of Alexandrine Parakeets as Nanday Parakeets was further reduced by 1.6%; Blue-crowned Parakeets as Alexandrine Parakeets by 1.0%; Blue-crowned Parakeets as Nanday Parakeets

Table 6.10: The confusion matrix based on the **Random Forest** classifier with the **selected combined features (top 70 features)**. The results are based on the correlation-based technique.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	84.6%	5.0%	1.3%	0.3%	0.3%	1.4%	0.7%	0.9%	0.2%	1.7%	1.2%	2.3%	0.3%	85%	7845
b=Nanday Parakeet	7.8%	84.2%	1.6%	1.0%	0.4%	1.1%	0.1%	0.5%	0.8%	0.5%	0.5%	0.6%	0.7%	84%	6246
c=Blue-crowned Parakeet	5.3%	2.6%	87.0%	0.4%	0.4%	0.9%	0.3%	0.9%	0.2%	0.4%	0.2%	1.0%	0.3%	87%	5332
d=Common House Martin	0.1%	0.2%	0.0%	98.5%	0.0%	0.0%	0.3%	0.0%	0.0%	0.4%	0.1%	0.0%	0.4%	98%	17896
e=Eastern Rosella	1.4%	1.6%	1.3%	0.6%	81.0%	1.0%	0.9%	1.0%	0.0%	5.0%	1.5%	4.3%	0.2%	81%	3247
f=Budgerigar (yellow)	3.1%	2.5%	0.6%	0.6%	0.2%	86.1%	1.1%	1.1%	0.1%	2.2%	1.0%	1.1%	0.3%	86%	4329
g=House Sparrow	0.3%	0.1%	0.0%	12.6%	0.2%	1.0%	71.5%	0.4%	0.0%	4.0%	1.1%	1.0%	7.8%	71%	5318
h=Budgerigar (wild-type)	8.5%	4.1%	3.9%	1.6%	0.7%	4.8%	2.6%	62.4%	0.9%	4.7%	1.8%	3.2%	0.7%	62%	3349
i=Common Wood Pigeon	2.6%	8.9%	3.7%	3.9%	0.7%	1.0%	0.7%	0.7%	67.2%	1.7%	3.2%	1.1%	4.4%	67%	2027
j=Black-headed Gull	0.1%	0.0%	0.0%	0.4%	0.1%	0.0%	0.1%	0.0%	0.0%	98.9%	0.1%	0.3%	0.0%	99%	29695
k=Cockatiel	0.5%	0.5%	0.0%	4.6%	0.2%	0.4%	1.2%	0.1%	0.1%	3.5%	84.8%	0.7%	3.3%	85%	5687
l=Budgerigar (blue)	2.0%	0.7%	0.2%	1.3%	0.8%	0.5%	0.9%	0.4%	0.0%	12.1%	1.4%	79.1%	0.6%	79%	7030
m=Common Starling	0.5%	0.7%	0.2%	8.4%	0.1%	0.7%	9.3%	0.1%	0.1%	0.9%	2.8%	0.4%	75.6%	76%	5392
Overall Correctly Classified														**89%**	

by 1.9%; and Nanday Parakeets as Blue-crowned Parakeets by 2.3%. This observation further validates the use of motion features to increasing correct classification rates, in species with closely related appearance.

Majority and Minority Sampled Species

On the thirteen classes dataset, the classification rate for the majority class (Black-headed Gull) is between 82% - 99% for all four standard classifiers and the minority class (Common Wood Pigeon) is between 22% - 67% (see Table 5.7). The minority class was not the class with the lowest classification rate but rather the Budgerigar (wild-type) class, with a classification rate of approximately 26% - 62% across all four classifiers. The confusion matrix Table 5.8 shows that only 1.7% of minority class (Common Wood Pigeon) were misclassified as majority class (Black-headed Gull). This was 4.7% for the Budgerigar (wild-type) class, which was the next class with lower samples after the Common Wood Pigeon class. They had the lowest correct classification rate as more of them (16.5%) were misclassified as Parakeets (8.5% as Alexandrine Parakeet, 4.1% as Nanday Parakeet and 3.9% as Blue-crowned Parakeet). The reason for this misclassification is that Budgerigar (wild-type) is more related in appearance (colour) to the Parakeet species. The Budgerigar classes were also with smaller samples and therefore misclassified as the majority class (Black-headed Gull). There were 4.7%, 12.1% and 2.2% Budgerigar (wild-type), Budgerigar (blue) and Budgerigar (yellow) respectively, that were misclassified as Black-headed Gull.

6.5.4 Contribution of the Motion Features

The previous section shows how feature selection can increase the correct classification rates when using the combined set of appearance and motion features. Table 6.4 also shows the results obtained using the 49 appearance features only (that is, with the motion features removed from the subset of 70). Comparing the two sets of figures allows us to estimate the contribution of the motion features to the overall classification rate.

Inspection of the result shows that there is an increase of 4% and 9% in the correct classification rates for the RF and RT classifiers respectively, when the combined features are used compared with when only appearance features are used. The primary interest is in the RF classifier, which is consistently best performing: in this case a 4% increase in classification rates, from 85 to 89% is significant. The statistical significance of the correct classification rate of combined set with feature selection was compared with that of appearnce with feature selection by computing the Wilcoxon's test statistic $W = 0$ (the smaller of $W^+ = 665$ and $W^- = 0$). Since the computed test statistic is less than 415, we accept the alternative hypothesis that the correct classification rate of the combined set with feature selection is significantly different from that of the appearance set with feature selection. Hence, the assertion that the 4% increase in classification rates is significant.

Comparing the appearance feature subset in Table 6.4 with the complete appearance feature set in Table 4.8 further supports this conclusion. There is a small observed difference when feature selection is used in conjunction with the appearance features alone; similarly, simply adding the motion features naively to produce a large set of 320 features similarly makes little impact. However, when feature selection is used in conjunction with the full set of motion and appearance features, a significant increase in correct classification rates is evident.

6.6 Conclusion

In this chapter, the work on automated classification of bird species in flight, using combined features (motion and appearance) have been further addressed. This chapter has addressed the challenge as a robust fine-grained classification problem using the combined

features and has shown experimentally that motion features are important for classification of species especially those with fine grain appearances.

Classification rates dropped significantly by approximately 9% for some of the classifiers when the combined features with the thirteen classes were used. Two feature selection techniques were then applied to select the feature set that produces the best classification results. The best classification rates for both methods occurred at 70 features. The classification rate was 75.40% for the correlation-based (CoBfs) method and 74.82% for the classifier-based. This was 49 selected optimal appearance features and 21 motion features with the correlation-based, and 62 appearances and 8 motion with the classifier-based. Finally, we noted that wing beat frequency and vicinity features were selected irrespective of the method used. This shows how important these groups of features were in classifying species by motion.

Experiments performed in Chapters 4 - 5 were revisited, using the selected features by the correlation-based method as they produced the best correct classification rates. The results using the selected features show that the performance of the RF and RT classifiers were superior to both NB and SVM. The classification results from the selected feature set were compared with those without feature selection.

This showed an increase in the correct classification rates by between 0-7% when RF and RT classifiers are used with the motion features, between 4-9% with the combined features, and 0-5% with appearance features. Surprisingly, misclassification of species with closely related appearances decreased with the selected appearance, motion and combined features. Specifically, with the selected combined features, there was between 1.0 - 2.3% reduction in misclassification, between 0.1% - 2.5% reduction in misclassification when appearance features were used and 0% - 0.8% when motion features were used.

The contribution of the selected motion features to the overall performance of the classifiers were also evaluated. There was an increase of 4% and 9% in correct classification rates for the RF and RT classifiers respectively. The best-performing classifier (RF) improved the classification rate by approximately 4%, which may be directly attributed to the use of motion features. Further analysis also revealed specific improvements in

species with similar visual appearance.

The works in Chapters 4 - 6 present results based on classification using single frames and subsets of videos. In the following Chapter, this will be extended to combine the results of several frames from a sequence using majority voting with the four standard classifiers in an attempt to improve classification rates.

Chapter 7

Improving the Performance of Our Bird Species Classifiers

The works presented in the previous chapters were based on results of classification using single frames and subsets of videos. In Chapter 5, species were classified using a combination of appearance and motion features in order to improve classification rates. However, the naive addition of motion to the appearance features only led to a small or no improvement to the classification rates. Irrelevant and or redundant features were then eliminated in Chapter 6 by performing feature selection in an attempt to further improve classification rates. Two features selection methods were used, namely correlation- and classifier-based techniques, which improved the correct classification rates by approximately 4%.

Most recent classifiers like the random forest classifier have novel classification results since they apply some sort of voting schemes (majority votes), which have greatly motivated the work presented in this chapter. In addition, majority voting technique have been successfully used by Bhattacharya and Chaudhuri (2003) to improve overall classification rates by combining the output of several classifiers. Since this research tracked and classified flying species in a video sequence frame by frame, it is beleived that aggregating these results will further improve the classification rate as previously done in Marini et al. (2013) for still images.

The aim of work presented in this chapter is to attempt a further increase in the correct classification rates. This is achieved by using majority voting techniques to aggregate the classification results presented in Chapter 6 across a set of video sub-sequences. This technique is applied to both the seven species and thirteen classes dataset using all four standard classifiers and the results were present. The rest of this chapter is structured into the following sections:

- In Section 7.1 the datasets used were presented, and the processing methods applied to the datasets in order to extract features and process sequences were described.
- Experimental work on classification using video sequences and majority voting were also described in Section 7.2 and the results presented in Section **??**
- Finally, conclusions were drawn to the results of the majority voting techniques in Sections 7.4

7.1 Dataset and Features Extraction

Datasets #2, which was described in detail in Chapter 4 was used for experiments presented in this chapter. As a reminder, these is the extended set of videos covering thirteen classes made up of eleven bird species with the Budgerigar (*Melopsittacus Undulatus*) having three colour forms. Specifically, it has been used in this chapter to classify bird species, by aggregating the results of several frames from a video sequence using majority voting.

The background Gaussian mixture model proposed by Zivkovic and van der Heijden (2006) was used to extract birds' silhouettes from each video. Contours were then obtained using the algorithm proposed by Suzuki et al. (1985) and were used to form connected components. Details of these have been presented in Chapter 3. The same techniques in Chapter 4 to extract appearance features. This was done by fitting oriented bounding box to each silhouette and extracting metrics like the height, width and hypotenuse, centroid, contour points and the silhouette itself. Appearance features made up of colour moments, shape moments, greyscale histogram, Gabor filter and log-polar

were then extracted from these metrics. Finally, centroids of the oriented bounding box were used to form trajectories (details in Chapter 5) from which the motion features were extracted. Motion features extracted from these trajectories include curvature scale space (CSS), turn-based, wing-beat frequency, centroid distance function (CDF), vicinity and curvature based on sine and cosine.

The appearance and motion features were merged to form the combined feature set (see Chapter 5), which were optimised using a correlation- and classifier-based feature selection technique described in Chapter 6. The optimally selected features by the classifier-based techniques were used to perform majority voting experiments presented in this chapter, as they yield the best results when compared with the correlation-based method. Again as a reminder, features used in this chapter have been represented as statistics, in order to reduce the feature dimension and enable real-time classification. The statistical features computed include the mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings.

7.2 Majority Voting Experiments

The experiments performed in this section aggregate the classification results from the extended dataset using majority voting. The experiments aggregate the classification results from the selected optimal features presented in Chapter 6, using majority voting technique and the results based on the four standard classifiers presented in Section 7.3.

Again, a five-fold cross-validation scheme was used, in which the complete dataset was randomly partitioned into five equal-sized subsets. Five iterations were then performed: for each iteration a unique subset was retained as a validation set and the others used for training. The results from each fold were then averaged to obtain the results for each experiment. The setup used for these experiments and classifier settings were the same as described in Chapters 4 and 5. To classify video sequences, the bird's appearance and motion features were first concatenated, as described in Chapter 5. In order to determine which class a bird in a video sequence belongs to, majority voting was applied to aggregate the results of each frame in the video sequence.

Suppose for a certain classification problems there are different classification outcomes for the j^{th} class based on the results of each individual image frame of a video sequence, $C_{1,j}, C_{2,j} \ldots C_{n,j}$. The n outcomes can be combined in such a way as to produce a more superior classifier. A common way to combine this outcome is expressed in Equation 7.1

$$Class_j = mode\{C_{1,j}, C_{2,j} \ldots C_{n,j}\} \tag{7.1}$$

In order words, each value of j is classified to the class that receives the largest number of classification (or votes). This has been known in literature by James (1998) as a majority vote classifier or majority vote learner. In general, the majority vote classifier consisting of votes from rules $C_1, C_2 \ldots C_n$ can defined in Equation 7.2.

$$Class_j = \arg\max_j \sum_{i=1}^{n} C_{i,j} \tag{7.2}$$

Where $C_{i,j}$ is the decision of the classifier at frame i for class j (which is 1 if class j is voted and 0 otherwise) and $Class_j$ is the decision of the classifier for the entire video sequence.

7.3 Majority Voting Results

The aim of this section is to determine if the aggregated results using majority voting outperform the results presented in Chapter 6, which uses the selected feature set. Table 7.1 shows results with and without majority voting using the extended set of thirteen classes across the four standard classifiers with the optimal feature set.

Firstly, based on results in this chapter, it is still evident that the performance of the two classifiers (RF and RT) are superior to both NB and SVM. Again, the RF classifier gives the highest correct classification rate with and without majority voting and based on the optimal feature set, thus the assertion that this is the most effective classifier of the four tested for this problem domain. In particular, the RF classifier outperforms SVM by up to 41% across all species. The Random Tree classifier similarly had a better performance

Table 7.1: Summary of species correct classification rates using the selected optimal features by the **classifier-based technique** in Chapter 6 **with** and **without** majority voting.

With Majority Voting				
Species	**NB**	**RF**	**RT**	**SVM**
a=Alexandrine Parakeet	96% ± 0.75%	97% ± 0.34%	97% ± 0.15%	99% ± 0.42%
b=Nanday Parakeet	97% ± 0.21%	97% ± 0.87%	98% ± 0.51%	93% ± 0.39%
c=Blue-crowned Parakeet	97% ± 0.58%	95% ± 0.52%	97% ± 0.48%	98% ± 0.22%
d=Common House Martin	99% ± 0.21%	100% ± 0.86%	97% ± 0.76%	97% ± 0.68%
e=Eastern Rosella	93% ± 0.62%	100% ± 0.66%	95% ± 0.43%	90% ± 0.44%
f=Budgerigar (yellow)	96% ± 0.56%	100% ± 0.02%	96% ± 0.53%	96% ± 0.33%
g=House Sparrow	97% ± 0.51%	99% ± 0.54%	99% ± 0.62%	89% ± 0.29%
h=Budgerigar (wild-type)	85% ± 0.76%	100% ± 0.38%	98% ± 0.28%	59% ± 0.49%
i=Common Wood Pigeon	89% ± 0.35%	91% ± 0.23%	89% ± 0.66%	57% ± 0.69%
j=Black-headed Gull	100% ± 0.39%	100% ± 0.26%	100% ± 0.38%	100% ± 0.48%
k=Cockatiel	98% ± 0.67%	100% ± 0.24%	97% ± 0.63%	98% ± 0.51%
l=Budgerigar (blue)	95% ± 0.43%	99% ± 0.41%	97% ± 0.43%	79% ± 0.68%
m=Common Starling	94% ± 0.69%	100% ± 0.48%	94% ± 0.80%	97% ± 0.62%
Overall Correctly Classified	**95% ± 0.52%**	**98% ± 0.49%**	**96% ± 0.44%**	**89% ± 0.42%**
Without Majority Voting				
a=Alexandrine Parakeet	39% ± 0.65%	85% ± 0.63%	64% ± 0.45%	56% ± 0.08%
b=Nanday Parakeet	53% ± 0.17%	84% ± 0.35%	61% ± 0.43%	44% ± 0.19%
c=Blue-crowned Parakeet	58% ± 0.24%	87% ± 0.65%	69% ± 0.47%	52% ± 0.71%
d=Common House Martin	91% ± 0.87%	98% ± 0.27%	89% ± 0.86%	97% ± 0.35%
e=Eastern Rosella	57% ± 0.18%	81% ± 0.83%	59% ± 0.02%	34% ± 0.49%
f=Budgerigar (yellow)	72% ± 0.61%	86% ± 0.72%	63% ± 0.54%	73% ± 0.31%
g=House Sparrow	57% ± 0.82%	71% ± 0.86%	53% ± 0.44%	40% ± 0.65%
h=Budgerigar (wild-type)	31% ± 0.36%	62% ± 0.29%	44% ± 0.71%	26% ± 0.87%
i=Common Wood Pigeon	50% ± 0.87%	67% ± 0.08%	51% ± 0.19%	22% ± 0.79%
j=Black-headed Gull	82% ± 0.28%	99% ± 0.53%	92% ± 0.53%	96% ± 0.13%
k=Cockatiel	54% ± 0.29%	85% ± 0.42%	65% ± 0.27%	58% ± 0.71%
l=Budgerigar (blue)	57% ± 0.69%	79% ± 0.66%	63% ± 0.15%	41% ± 0.73%
m=Common Starling	66% ± 0.77%	76% ± 0.56%	57% ± 0.22%	72% ± 0.63%
Overall Correctly Classified	**68% ± 0.57%**	**89% ± 0.52%**	**74% ± 0.33%**	**71% ± 0.41%**

than either NB or SVM on these selected optimal features. To confirm that the Random Forest (RF) classifier's correct classification rate is statistically significant compared with the other classifiers, the Wilcoxon's test statistic W is calculated. For RF compared with NB classifiers, $W = 0$ (the smaller of $W^+ = 679$ and $W^- = 0$), when RF and RT are compared, $W = 0$ (this is the smaller of $W^+ = 665.5$ and $W^- = 0$) and $W = 0$ (the smaller of $W^+ = 623$ and $W^- = 0$) when RF is compared with the SVM classifier. Since the NB classifier compared with the RF classifier resulted in a W less than 415, we accept the alternative hypothesis that the correct classification rate of the RF classifier is significantly different from that of the NB classifier. This is also true for the RF compared with the RT and SVM classifiers, as both resulted in a W which is less than 415. Hence, the assertion that the RF classifier gives the highest correct classification rate, which is statistically significant from those of the other three classifiers when the seven species dataset is used with the appearance feature sets.

The results of majority voting also show an improvement when compared with those without majority voting across all the four standard classifiers. In particular, based on the best performing classifier, the majority voting showed 9% increase in correct classification results. There was also an increase of approximately 22%, 27% and 18% when the RT, NB and SVM classifiers were used respectively. This shows that aggregating the classification results using majority voting improves the results.

Appearance-Related Species

The fine-grained performance of the majority voting method were examined; that is, differentiation between species of similar appearance. Detailed cross-species confusion matrix for the RF classifier obtained using the optimal features with majority voting have been presented in Table 7.2. The cross-species confusion matrix for the other classifiers have been presented in Appendix G. When classification results were aggregated with majority voting, there was less misclassification among these species than when no aggregation technique was applied. For example, the misclassification of Alexandrine parakeets as Nanday parakeets reduced from 5.0% to 1.3%; Nanday parakeets as Alexandrine parakeets from 7.8% to 0%; and Blue-crowned parakeets as Alexandrine parakeets from

Table 7.2: The confusion matrix of video classification based on the **Random Forest** classifier with the **selected optimal combined features (70 features)** by the **classifier-based technique**.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Videos
a=Alexandrine Parakeet	97.4%	1.3%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	77
b=Nanday Parakeet	0.0%	96.6%	3.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	59
c=Blue-crowned Parakeet	1.7%	0.0%	94.8%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.0%	0.0%	95%	58
d=Common House Martin	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%	114
e=Eastern Rosella	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%	40
f=Budgerigar (yellow)	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%	47
g=House Sparrow	0.0%	0.0%	0.0%	1.4%	0.0%	0.0%	98.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	99%	74
h=Budgerigar (wild-type)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%	41
i=Common Wood Pigeon	0.0%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	2.9%	91.4%	0%	0.0%	0.0%	0.0%	91%	35
j=Black-headed Gull	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100%	142
k=Cockatiel	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100%	58
l=Budgerigar (blue)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	98.7%	0.0%	99%	76
m=Common Starling	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100%	71
Overall Correctly Classified														**98%**	**892**

5.3% to 1.7%. This suggests that aggregating classification results using majority voting may increase the correct classification rates for species with closely-related appearances.

Majority and Minority Sampled Species

The classification rate for the majority class (Black-headed Gull) is 100% for all four standard classifiers and the minority class (Common Wood Pigeon) is between 57% - 91% (see Table 7.1). The minority class is the class with the lowest classification rate, with the classification rate of approximately 57% - 91% across all four classifiers. The confusion matrix Table 7.2 shows that 0% of minority class (Common Wood Pigeon) were misclassified as majority class (Black-headed Gull), which was 12.9% when no majority voting nor feature selection was applied and 1.7% when only feature selection was applied. This may suggest that the introduction of majority voting, may partly resolve the imbalance problem with the extended dataset.

7.4 Conclusion

In this chapter, experimental results of the combined features with majority voting technique to support ecological studies or migration and other population-level behaviours have been presented with excellent results. The majority voting technique was used to perform experiments that aggregated classification results using the smaller dataset of seven species and the selected optimal features based on the larger dataset in Chapter 6. The classification of birds in flight from a video is particularly challenging for automated species identification and no existing work has yet addressed this problem directly.

The RF classifier had produced consistently the highest correct classification rate based on all experiments performed in this thesis. In this chapter, it has produced the highest classification rates among all the classifiers used based on the selected combined feature set with and without majority voting. Thus, the assertion that this is the most effective classifier of the four tested for this problem domain.

When the majority voting results were compared with those without, there was an increase in correct classification rates across all four classifiers. This was approximately 9%, 22%, 27% and 18% increase when the RF, RT, NB and SVM classifiers were used respectively. Thus, showing that aggregating the classification results may improve correct classification rates.

Finally, the aggregation results showed that there was less misclassification among species with related appearances compared with when no aggregation was applied.The results showed that Alexandrine parakeets misclassified as Nanday parakeets reduced by 3.7%; Nanday parakeets as Alexandrine parakeets reduced by 7.8% and Blue-crowned parakeets as Alexandrine parakeets reduced by 3.6%. Thus suggesting that aggregating classification results using majority voting may further increase the correct classification rates of these species.

Various attempts have been made to improve classification rates of the various models used in this thesis. Even though the results have improved significantly with these methods, it is recommended that, for future work, imbalanced dataset techniques be further investigated with the extended datasets. It is believed and has been shown in subsequent chapters that balancing the dataset has a likelihood to further improve the classification accuracies of the system.

Chapter 8

Conclusions

This thesis has explored some novel techniques for the classification of bird species using computer vision and machine learning. Appearance features from previous related works were evaluated and refined for use in automatic classification of species in flight. Relevant motion features which can be extracted from video of birds in flight were also identified and used to classify species automatically. The appearance and motion features were then effectively combined to determine if they give better results than motion or appearance features alone. Redundant and irrelevant features were then eliminated from the combined feature set to determine if they can further improve classification accuracy. Finally, a further refinement of this technique, using majority voting to aggregate classification results across a set of video sub-sequences was undertaken. A video dataset of birds in flights for use by the research community have also been collated. The following original contributions have been presented in the preceding sections:

1. Fast Fourier Transform (FFT) has been used for the measurement of wing beat frequencies. This has been initially evaluated on bat species and later on bird species. These methods have been evaluated with the state-of-the-art by Cutler and Davis (2000) and have been found to outperform it significantly.

2. A new set of appearance features have also been developed for bird species classification. This shows an improvement in correct classification rate over the state-of-the-art proposed by Marini et al. (2013) on the three datasets introduced in this thesis.

3. A new set of motion features were developed but using these alone could not effectively classify bird species. As a result, a species classification scheme was then developed, which combines motion and appearance features with the aim of improving classification accuracy. This is the first work that combines features in this way and it achieved a correct classification rate of 85% which outperformed the existing state of art method.

4. The naive combination of appearance and motion features did not significantly improve classification accuracy but rather introduced redundant and irrelevant features. Two feature reduction techniques were therefore evaluated against the combined set, which increased classification accuracy by approximately 4% when the RF classifier was used.

5. Majority voting technique was used to aggregate the classification results of the selected optimal features, obtained with classifier-based feature selection technique, across an entire video sequence. Thus increasing the classification rate to between 89% and 98%.

6. A video dataset was collated against which future work can be benchmarked. This thesis resulted initially in a dataset of seven species and later in an extended dataset of thirteen classes consisting of eleven species, one of which had three colour forms.

8.1 Main Contributions, Limitations and Future Work

In this section, the contributions mentioned above have been summarised, together with a discussion of their limitations if any and possible future work.

8.1.1 Analysis of Wing Beat Frequency

Periodic motion features associated with wing beats vary among species, and this was proposed to provide an additional useful discriminating feature for classification of bird species. Two modified techniques based on bounded box metrics and similarity matrices were proposed for the analysis of wing beat frequency. A preliminary investigation using a small dataset of bats in flight showed that these methods significantly outperformed

the baseline algorithm, especially when the bat's orientation changed. The method based on self-similarity matrices was found to be better than the baseline algorithm when the orientation of bats do not change. Where there is one obvious peak above the set threshold value, all methods are able to estimate the correct wing beat. The baseline algorithm usually considered this as an ideal situation and rejects all others as non-ideal. However, the proposed methods were able to estimate a frequency closer to the ground truth when there is more than one peak above the set threshold or when all peaks are below the threshold.

The major challenge working with bat dataset is that they are usually filmed at night and the image quality is low. The wing beat frequency measurements were based on fitting a bounding box on a silhouette of birds and taking some metrics. Therefore low video quality directly affects the extracted metrics and hence the wing beats frequencies estimated.

It was observed that larger window size may help improve the frequency resolution, thus improving the accuracy of the wingbeat frequencies. This can be achieved by either increasing the time domain sample and/or using spectral interpolation (Lyons, 2010). The first is usually ideal but this depends on the collected data, which in this case was made up of shorter sequences due to the narrow camera's angle of view. For future work using a 360-degree camera is recommended to capture longer sequences and therefore larger window sizes, as they offer the widest field-of-view available, without blind spots.

8.1.2 Classification using Appearance Features

Using image-based (appearance) features to classify species provided a richer and more informative data for classification of species. A new set of appearances features was developed for the classification of species, which combines colour, shape and texture. Specifically, colour moments and colour log-polar, shape moments, Gabor filters and greyscale histograms features were used, which were extracted from individual images and concatenated to form the appearance feature set for classification using four standard classifiers.

The combined appearance features (colour, shape and texture) outperforms methods

that use only colour features. Particularly, when the appearance features were compared with the state-of-the-art that used colour, it achieved a better classification performance. A better performance was also achieved with the RF classifier than the other classifiers when used with both the seven species and thirteen classes datasets. Classification rates dropped significantly when moving from seven to thirteen classes. When the number of classes was increased from 2 to 5, 17 and 200, the classification rates dropped with both the proposed methods and the state-of-the-art method that uses colour features alone. Even though performance dropped when the number of classes was increased, the classification rates of the appearance features were observed to be higher than the state-of-the-art on the Caltech-UCSD Birds-200-2011 dataset. It was noted that increasing the number of classes whiles using appearance features alone may result in a decrease in classification rates irrespective of appearance features used. Finally, experimental results revealed that misclassification of bird species was due to illumination and similar colour patterns in some species. The state-of-the-art method which used only colour features had more misclassifications with similar colour patterned species than the appearance feature set.

The identification of a suitable set of appearance features is very challenging. Most existing works use appearance-based (mainly colour) features in conjunction with single images of static birds. Whilst these existing works are a very useful starting point, there is no comparable work for in-flight birds. A set of appearance features has been provided for use with birds in flight. Whilst this work has been shown to work effectively with seven and thirteen classes, there is also a scope for future work to explore an acceptable feature set and techniques for classifying images of non-static birds.

When the number of classes is increased in the dataset, the classification rates can be improved (classifying with appearance features) by applying part localisation techniques (Gavves et al., 2013; Krause et al., 2015). Semantic part localisation can facilitate fine-grained categorisation by explicitly isolating subtle appearance differences associated with specific object parts. However, most techniques that use this approach manually mark birds parts. In this research, manually marking parts in a second of video will involve marking these in 240 image frames. For future work, it is recommended that an

annotation-free based technique (Gavves et al., 2013; Krause et al., 2015; Gavves et al., 2015) be developed to localise bird parts using co-segmentation and alignment before exacting the appearance features, as these will increase the classification rates.

8.1.3 Classification using Appearance and Motion features

The appearance-based feature set used single images and appearance-based models for classification. However, bird species also exhibited distinguishing behaviours (flying, and poses, etc) and this was used together with appearance features to improve classification performance, especially at a distance where appearance-based features such as colour attenuated. A robust system that combined the appearance and motion features for classification of bird species was developed. This system fused the appearance-based features in the previous section with motion cues, such as CSS, turn based, wing-beat frequency, centroid distance function, vicinity and curvature based on sine and cosine to improve classification performance. All experiments performed with this feature set used the four standard classifiers.

This research revealed the RF classifier to have the best classification rate across all datasets when the combined feature set was used. Motion cues are also the most effective for classifying species at a distance, where coloured features attenuate. Finally, naively fusing motion and appearance features introduced redundant features, and thus made little contribution to the classifiers performances (classification rates).

The identification of suitable motion features remains challenging. Even though these have been explored for the classification of fish trajectory and human motion identification, there is no significant work which considers bird species. A set of combined features (appearance and motion) have been provided for use with birds in flight. Whilst this work has been shown to work with seven and thirteen classes, there was a small increase in the correct classification rates when compared with appearance. However, combining features resulted in a larger feature set, which introduced redundant and irrelevant ones. It was hypothesised that selecting the most optimal feature set from the combined set can help improve classification rates. There is still a scope for further work to explore an acceptable motion feature set and techniques for classifying images of non-static birds.

Flight style can be a great way to identify birds at a distance. Motion features were used to improve classification of species with closely related appearances. The most important motion features that help in differentiating species are wing beat frequencies. However, these are affected by the temporal resolution used in the calculation of the wing beats. Longer tracks will lead to higher temporal resolution and therefore improve the classification of species using motion. For future work the use of a 360-degree camera is therefore, recommended in order to capture longer sequences and improve the frequency resolution.

8.1.4 Classification using the Optimal Feature Set

It is been hypothesised that combining appearance and motion features may improve classification performance. However, the combined results without feature selection had a small increase in the correct classification rates. It was hypothesised that performance of the combined set may be undermined by the presence of redundant features and that significantly better classification rates may be achieved using a subset of the full combined set. Two feature selection techniques have been explored, which helped in the selection of optimal features from the combined set to improve correct classification rates.

The two feature selection methods both showed that the majority of optimal features selected were appearance features. However, some motion features (wing beat frequency and vicinity) were also selected irrespective of the method used, which suggest their importance for classification. The main finding was that the addition of motion features with feature selection, significantly improved classification rates and a further analysis also revealed specific improvements in classification rates for species with similar visual appearance.

8.1.5 Improving Classification using Other Techniques

Again, in an attempt to increase classification rates of the proposed models, experiments were performed that aggregate the classification results of the selected optimal features by the classifier-based method using the majority voting technique.

Aggregating the results of the combined features with majority voting outperforms

those without on both the seven species and thirteen classes datasets. The results also showed that the RF and RT classifiers had better performances than either NB or SVM. Finally, the misclassification among bird species with closely related appearances improved when majority voting technique was applied.

Motion features have been extracted from a sequence of image frames whilst appearance features were extracted from individual images. Now, determining an effective way to combine these features for classification in the absence of comparable existing work is challenging. Therefore, even though some initial work in this area have been presented, there is still scope for further studies to explore other ways of combining this problem.

8.2 Application of this Research

The collection of species population data is key to scientists in determining whether to build offshore wind turbines and ecologist; for managing farmland use and studying migration behaviour but all these requires automated censusing of large populations. There has been some work (Cullinan et al., 2015; Matzner et al., 2015; Hristov et al., 2010; Betke et al., 2008; Lazarevic et al., 2008) that uses computer vision techniques to automated censusing of large bat populations but these were particularly not for species-level censusing. Most of these have used a thermal camera with computer vision algorithms to only detect and track species. The work presented in this thesis has not only been used to detected and track bird species but has also been able to classify them using their motion and appearance features, which is believe will be essential in collecting behaviour data of species for ecological and biological studies.

In order to conserve species, there is the need to establish the size of current populations (censusing), working out which species are where and how they are responding to the threats and pressures they face and also identifying which species need action now and what threats they face. Species-level identification and censusing have not been studied and the work undertaken can be applied to this problem since the system presented in this thesis is able to identify species from their characteristic flight pattern and appearances. Another important area that biologist wish to understand is the behaviour of different

species, including migratory behaviour, this work again can be applied to these and most importantly can be used to identify migratory species.

The classification of bird species using either motion and/or appearance features can be treated as a fine-grained problem, which is similar to the classification of bats, butterflies and fish. Hitherto, no work has used a combination of appearance and motion features to classify these at species-level. Again, this work could be applied to classify these species and aid in collecting important biological and ecological data.

Little comparable work has addressed the problem of bird species classification and none has achieved robust identification by combining appearance and motion features; this work, therefore represents a significant contribution with potential benefit to ecological monitoring and data collection.

Appendix A

Appearance without Feature Selection

Table A.1: The confusion matrix based on the **Naive Bayes** classifier without feature selection, using **the appearance features** on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	27.2%	17.2%	6.1%	0.6%	1.2%	16.5%	1.2%	6.9%	2.7%	3.7%	6.2%	9.5%	1.0%	27%	12801
b=Nanday Parakeet	6.6%	46.6%	7.3%	1.2%	4.3%	6.9%	0.3%	1.3%	9.8%	2.2%	7.3%	5.7%	0.6%	47%	10025
c=Blue-crowned Parakeet	11.1%	10.7%	48.0%	1.5%	3.1%	10.9%	1.7%	5.5%	2.9%	1.0%	1.0%	2.4%	0.4%	48%	9076
d=Common House Martin	0.0%	0.7%	0.2%	77.2%	0.0%	0.0%	7.8%	0.1%	0.3%	0.0%	1.4%	0.1%	12.2%	77%	25517
e=Eastern Rosella	1.5%	9.2%	4.8%	1.9%	52.8%	4.4%	1.1%	2.7%	1.6%	2.8%	6.2%	10.8%	0.4%	53%	5929
f=Budgerigar (yellow)	3.2%	5.8%	4.9%	0.3%	0.7%	60.1%	2.1%	9.4%	2.2%	2.4%	7.2%	0.5%	1.3%	60%	7667
g=House Sparrow	1.7%	0.0%	1.6%	13.1%	0.6%	1.2%	44.6%	3.3%	0.0%	5.3%	4.9%	3.6%	20.1%	45%	10191
h=Budgerigar (wild-type)	4.1%	4.9%	10.3%	2.6%	4.0%	12.8%	6.1%	22.0%	5.0%	6.1%	11.4%	8.2%	2.5%	22%	6283
i=Common Wood Pigeon	3.0%	18.1%	2.9%	4.1%	2.1%	2.3%	1.5%	1.0%	47.3%	0.7%	7.4%	2.4%	7.2%	47%	4301
j=Black-headed Gull	2.8%	1.7%	0.5%	0.8%	1.1%	0.0%	7.0%	3.5%	1.5%	63.8%	8.4%	7.9%	1.0%	64%	38764
k=Cockatiel	1.2%	5.4%	0.3%	6.0%	1.8%	1.5%	3.0%	1.4%	1.1%	2.5%	55.5%	9.1%	11.2%	56%	9398
l=Budgerigar (blue)	1.4%	6.3%	2.2%	3.6%	8.3%	1.4%	6.0%	3.4%	2.1%	20.5%	8.9%	35.0%	0.9%	35%	12090
m=Common Starling	2.4%	1.2%	0.4%	15.0%	0.2%	2.9%	15.3%	0.4%	0.8%	1.3%	7.8%	0.1%	52.2%	52%	9865
Overall Correctly Classified														**54%**	**161907**

Table A.2: The confusion matrix based on the **Random Tree** classifier without feature selection, using **the appearance features** on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	61.2%	8.9%	5.9%	0.6%	2.1%	3.9%	1.2%	4.0%	1.7%	3.2%	1.6%	4.5%	1.1%	61%	12801
b=Nanday Parakeet	12.2%	56.3%	5.2%	1.4%	3.2%	4.1%	0.6%	3.3%	5.1%	1.7%	2.3%	3.2%	1.5%	56%	10025
c=Blue-crowned Parakeet	8.3%	5.8%	67.6%	0.4%	2.5%	3.0%	1.1%	3.8%	2.4%	1.0%	0.9%	2.1%	1.1%	68%	9076
d=Common House Martin	0.3%	0.6%	0.2%	85.7%	0.2%	0.2%	4.4%	0.4%	0.5%	1.1%	1.9%	0.6%	3.9%	86%	25517
e=Eastern Rosella	4.7%	5.5%	4.3%	0.6%	53.7%	2.1%	2.2%	3.8%	1.9%	6.1%	3.8%	10.2%	1.3%	54%	5929
f=Budgerigar (yellow)	8.1%	5.2%	4.3%	0.6%	2.2%	59.3%	2.6%	6.1%	1.6%	2.8%	2.2%	3.0%	2.0%	59%	7667
g=House Sparrow	1.5%	0.7%	1.1%	10.7%	1.3%	2.0%	51.9%	2.5%	0.8%	5.4%	3.3%	3.7%	15.1%	52%	10191
h=Budgerigar (wild-type)	9.7%	5.6%	6.0%	2.0%	3.8%	7.5%	4.4%	40.5%	2.5%	5.1%	3.3%	6.9%	2.8%	40%	6283
i=Common Wood Pigeon	5.3%	10.1%	4.3%	3.7%	2.3%	1.9%	1.2%	3.2%	55.6%	2.3%	3.6%	2.2%	4.3%	56%	4301
j=Black-headed Gull	1.2%	0.5%	0.2%	0.8%	1.0%	0.5%	1.6%	0.8%	0.3%	86.1%	1.6%	4.8%	0.6%	86%	38764
k=Cockatiel	2.4%	2.3%	0.8%	4.8%	2.0%	1.9%	3.9%	1.9%	1.6%	5.9%	60.9%	4.8%	6.8%	61%	9398
l=Budgerigar (blue)	5.4%	2.5%	1.7%	1.2%	4.6%	1.6%	3.1%	3.5%	1.0%	14.4%	3.6%	56.1%	1.3%	56%	12090
m=Common Starling	1.6%	1.8%	0.9%	10.7%	0.8%	1.4%	15.9%	1.5%	2.1%	2.1%	6.6%	1.6%	53.0%	53%	9865
Overall Correctly Classified														**68%**	**161907**

Table A.3: The confusion matrix based on the **SVM** classifier without feature selection, using **the appearance features** on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	56.4%	9.2%	7.0%	0.7%	0.4%	9.0%	0.7%	0.6%	1.5%	4.2%	2.5%	6.9%	0.8%	56%	12801
b=Nanday Parakeet	15.6%	47.0%	9.5%	1.4%	1.6%	5.9%	0.2%	0.5%	5.4%	2.0%	4.7%	5.1%	1.1%	47%	10025
c=Blue-crowned Parakeet	16.6%	6.4%	61.2%	0.9%	1.8%	5.2%	1.1%	0.7%	0.3%	0.8%	0.7%	3.8%	0.5%	61%	9076
d=Common House Martin	0.3%	0.3%	0.1%	95.8%	0.0%	0.0%	1.1%	0.1%	0.1%	0.4%	0.6%	0.1%	1.1%	96%	25517
e=Eastern Rosella	3.7%	4.8%	7.7%	1.1%	48.1%	1.7%	1.0%	0.5%	1.0%	9.3%	3.9%	15.1%	2.2%	48%	5929
f=Budgerigar (yellow)	7.1%	2.7%	3.9%	1.2%	1.0%	71.2%	1.3%	1.0%	0.5%	5.3%	3.0%	0.7%	1.2%	71%	7667
g=House Sparrow	0.5%	0.1%	0.6%	18.3%	1.5%	2.8%	39.0%	0.5%	0.3%	10.6%	2.6%	3.3%	19.9%	39%	10191
h=Budgerigar (wild-type)	12.7%	6.2%	8.7%	3.5%	3.8%	8.9%	5.8%	23.0%	2.8%	10.1%	6.4%	5.5%	2.6%	23%	6083
i=Common Wood Pigeon	6.5%	12.1%	4.6%	3.8%	1.9%	0.5%	0.6%	0.1%	53.1%	2.0%	4.3%	2.7%	7.9%	53%	4301
j=Black-headed Gull	0.5%	0.2%	0.1%	0.3%	0.1%	0.1%	0.7%	0.2%	0.1%	94.9%	0.6%	2.1%	0.0%	95%	38764
k=Cockatiel	1.9%	1.0%	0.2%	5.7%	1.2%	1.5%	0.8%	0.4%	0.6%	4.0%	62.9%	6.1%	13.7%	63%	9398
l=Budgerigar (blue)	5.7%	2.0%	1.6%	1.6%	3.4%	1.0%	2.8%	0.3%	0.6%	24.9%	6.5%	48.2%	1.4%	48%	12090
m=Common Starling	1.4%	0.2%	0.4%	10.5%	0.6%	1.0%	6.0%	0.1%	0.6%	1.4%	2.9%	0.8%	74.2%	74%	9865
Overall Correctly Classified														**70%**	**161707**

Appendix B

Motion without Feature Selection

Table B.1: The confusion matrix based on the **SVM** classifier without feature selection, using **the motion features** on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	20.7%	5.8%	0.5%	25.3%	0.0%	0.0%	1.6%	0.0%	0.0%	28.4%	2.8%	0.4%	14.5%	21%	7845
b=Nanday Parakeet	15.5%	11.1%	0.3%	28.8%	0.0%	0.0%	2.0%	0.0%	0.0%	20.5%	3.8%	0.1%	17.7%	11%	6246
c=Blue-crowned Parakeet	20.5%	9.5%	12.2%	20.0%	0.0%	0.0%	1.6%	0.0%	0.0%	13.8%	3.5%	0.3%	18.8%	12%	5332
d=Common House Martin	2.3%	0.6%	0.1%	36.6%	0.0%	0.0%	0.7%	0.0%	0.0%	57.3%	0.2%	0.2%	2.0%	37%	17896
e=Eastern Rosella	14.6%	3.5%	0.5%	25.4%	6.2%	0.0%	1.5%	0.0%	0.0%	36.1%	2.0%	0.2%	9.9%	6%	3247
f=Budgerigar (yellow)	5.9%	1.7%	0.2%	13.7%	0.0%	3.5%	1.4%	16.2%	0.0%	29.3%	0.3%	23.9%	4.0%	3%	4329
g=House Sparrow	8.4%	3.7%	0.2%	28.9%	0.0%	0.0%	8.7%	0.0%	0.0%	33.1%	0.5%	0.3%	16.1%	9%	5318
h=Budgerigar (wild-type)	10.7%	3.7%	0.1%	10.7%	0.0%	11.9%	3.9%	6.0%	0.0%	27.1%	0.6%	18.2%	7.0%	6%	3349
i=Common Wood Pigeon	31.9%	8.6%	0.9%	26.9%	0.0%	0.0%	2.5%	0.0%	4.9%	8.8%	2.0%	0.4%	13.1%	5%	2027
j=Black-headed Gull	4.2%	0.2%	0.0%	17.1%	0.0%	6.7%	0.1%	0.0%	0.0%	67.6%	3.6%	0.1%	0.3%	68%	29695
k=Cockatiel	6.2%	4.9%	0.2%	31.4%	0.0%	0.0%	0.7%	0.0%	0.0%	39.3%	6.9%	0.2%	10.2%	7%	5687
l=Budgerigar (blue)	5.4%	1.4%	0.1%	20.8%	0.0%	15.7%	1.1%	18.6%	0.0%	29.4%	0.8%	3.4%	3.3%	3%	7030
m=Common Starling	6.9%	4.9%	0.4%	22.9%	0.0%	0.0%	2.9%	0.0%	0.0%	13.6%	2.4%	0.2%	45.8%	46%	5392
Overall Correctly Classified														**33%**	**103393**

Table B.2: The confusion matrix based on the **Random Tree** classifier without feature selection, using **the motion features** on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	12.9%	8.7%	7.5%	14.9%	4.3%	4.8%	5.4%	4.1%	2.9%	14.1%	7.0%	6.6%	6.8%	13%	7845
b=Nanday Parakeet	11.9%	10.1%	8.4%	13.1%	4.0%	4.1%	6.2%	3.6%	3.1%	12.2%	8.2%	5.8%	9.5%	10%	6246
c=Blue-crowned Parakeet	12.1%	9.4%	9.0%	14.3%	4.2%	4.2%	6.0%	3.9%	2.6%	11.3%	8.2%	6.1%	8.8%	9%	5332
d=Common House Martin	6.4%	4.8%	3.9%	26.3%	3.0%	4.7%	4.1%	3.4%	1.5%	25.5%	4.7%	7.8%	3.7%	26%	17896
e=Eastern Rosella	10.1%	8.7%	5.5%	16.0%	5.2%	4.4%	5.7%	3.2%	2.1%	19.5%	6.8%	7.0%	6.0%	5%	3247
f=Budgerigar (yellow)	8.2%	5.2%	5.0%	10.6%	3.4%	7.3%	5.1%	15.5%	1.6%	12.7%	5.4%	15.5%	4.6%	7%	4329
g=House Sparrow	8.5%	6.8%	6.3%	14.8%	3.4%	4.2%	11.2%	4.1%	2.8%	16.9%	5.8%	7.2%	8.1%	11%	5318
h=Budgerigar (wild-type)	7.8%	5.6%	4.8%	7.6%	3.1%	14.6%	6.3%	5.8%	1.8%	14.0%	5.4%	17.4%	5.9%	6%	3349
i=Common Wood Pigeon	9.7%	9.8%	7.9%	12.7%	4.4%	4.6%	6.8%	4.5%	5.2%	10.1%	7.7%	7.7%	8.9%	5%	2027
j=Black-headed Gull	3.9%	2.3%	1.8%	15.8%	2.1%	3.5%	3.1%	2.5%	0.8%	53.0%	3.5%	6.0%	1.7%	53%	29695
k=Cockatiel	9.8%	7.8%	6.4%	15.1%	4.1%	5.0%	5.1%	3.3%	2.1%	18.4%	9.8%	6.7%	6.3%	10%	5687
l=Budgerigar (blue)	6.8%	5.2%	4.3%	11.7%	3.1%	12.9%	4.7%	12.5%	1.8%	15.5%	5.1%	12.3%	4.0%	12%	7030
m=Common Starling	10.4%	9.4%	8.3%	11.7%	4.1%	3.8%	8.2%	4.0%	3.3%	8.7%	7.1%	5.7%	15.4%	15%	5392
Overall Correctly Classified														**25%**	**103393**

Table B.3: The confusion matrix based on the **Naive Bayes** classifier without feature selection, using **the motion features** on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	2.9%	4.3%	6.2%	6.3%	12.0%	11.3%	0.8%	1.2%	4.9%	21.4%	17.7%	8.2%	2.9%	3%	7845
b=Nanday Parakeet	1.4%	6.6%	4.4%	5.8%	10.3%	12.2%	1.7%	0.8%	4.7%	14.0%	24.8%	8.5%	4.8%	7%	6246
c=Blue-crowned Parakeet	1.9%	3.8%	9.6%	5.4%	11.9%	12.0%	0.4%	1.3%	1.5%	12.1%	27.4%	9.6%	3.1%	10%	5332
d=Common House Martin	0.4%	0.6%	2.3%	13.1%	10.6%	4.5%	0.3%	0.5%	0.4%	60.4%	2.3%	4.0%	0.6%	13%	17896
e=Eastern Rosella	0.3%	1.9%	7.3%	6.1%	18.9%	10.9%	0.2%	0.5%	1.7%	30.2%	11.6%	9.4%	1.0%	19%	3247
f=Budgerigar (yellow)	0.4%	1.1%	2.0%	7.6%	12.3%	11.3%	0.7%	12.1%	0.9%	26.7%	4.1%	19.0%	1.9%	11%	4329
g=House Sparrow	1.6%	3.1%	6.1%	9.2%	17.0%	7.9%	5.3%	1.0%	0.9%	29.8%	4.8%	7.1%	6.0%	5%	5318
h=Budgerigar (wild-type)	1.0%	1.7%	4.5%	6.7%	11.9%	24.7%	1.8%	0.6%	1.1%	16.4%	6.8%	21.1%	1.6%	1%	3349
i=Common Wood Pigeon	5.0%	8.2%	5.8%	2.4%	11.8%	7.7%	1.6%	1.7%	16.7%	6.9%	18.6%	6.0%	7.6%	17%	2027
j=Black-headed Gull	0.2%	0.5%	0.4%	2.7%	5.7%	1.5%	0.4%	0.3%	0.6%	84.9%	1.4%	1.1%	0.4%	85%	29695
k=Cockatiel	0.5%	1.5%	1.4%	7.9%	10.0%	10.0%	0.5%	0.2%	0.9%	32.2%	24.8%	7.5%	2.7%	25%	5687
l=Budgerigar (blue)	0.9%	1.4%	2.3%	8.5%	10.8%	19.4%	0.9%	14.8%	0.9%	29.1%	3.5%	6.0%	1.4%	6%	7030
m=Common Starling	3.0%	2.7%	8.6%	7.6%	7.6%	11.9%	2.0%	0.2%	2.2%	11.0%	21.6%	8.8%	12.7%	13%	5392
Overall Correctly Classified														**32%**	**103393**

Appendix C

Combined without Feature Selection

Table C.1: The confusion matrix based on the **Naive Bayes** classifier without feature selection, using **the combined features** on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	39.7%	17.4%	6.4%	0.5%	2.0%	8.2%	0.8%	6.4%	3.1%	2.2%	6.3%	5.5%	1.3%	40%	7845
b=Nanday Parakeet	8.2%	53.2%	7.0%	0.5%	4.1%	3.8%	0.3%	1.1%	10.5%	0.6%	5.2%	4.6%	1.0%	53%	6246
c=Blue-crowned Parakeet	14.8%	10.5%	56.6%	0.8%	2.6%	3.4%	1.0%	3.8%	3.1%	0.8%	0.8%	1.1%	0.7%	57%	5332
d=Common House Martin	0.0%	0.8%	0.1%	89.5%	0.0%	0.0%	4.7%	0.2%	0.6%	0.4%	1.0%	0.1%	2.7%	90%	17896
e=Eastern Rosella	3.0%	9.3%	3.4%	2.2%	51.0%	4.4%	0.9%	1.8%	1.4%	5.9%	6.3%	10.2%	0.2%	51%	3247
f=Budgerigar (yellow)	1.9%	2.3%	1.5%	0.7%	0.3%	69.8%	1.6%	10.7%	1.1%	0.8%	4.3%	4.2%	0.9%	70%	4329
g=House Sparrow	1.1%	0.1%	0.9%	10.1%	0.4%	1.8%	59.3%	3.8%	0.1%	5.4%	2.9%	2.7%	11.4%	59%	5318
h=Budgerigar (wild-type)	5.0%	3.0%	4.4%	2.2%	3.0%	14.5%	2.1%	33.5%	4.3%	6.9%	10.3%	9.6%	1.2%	34%	3349
i=Common Wood Pigeon	2.1%	13.9%	3.4%	4.6%	1.6%	0.6%	0.7%	0.9%	55.0%	0.1%	8.6%	1.9%	6.6%	55%	2027
j=Black-headed Gull	1.9%	1.1%	0.5%	1.1%	1.2%	0.1%	1.0%	0.6%	0.1%	84.0%	4.3%	4.1%	0.0%	84%	29695
k=Cockatiel	1.7%	4.9%	0.5%	5.4%	2.8%	1.4%	2.3%	0.8%	1.6%	2.3%	63.1%	6.2%	7.2%	63%	5687
l=Budgerigar (blue)	2.2%	2.1%	1.2%	3.2%	3.6%	2.8%	3.1%	5.3%	1.3%	11.3%	7.6%	55.6%	0.6%	56%	7030
m=Common Starling	2.2%	1.1%	0.4%	8.3%	0.0%	1.9%	7.9%	0.3%	0.5%	1.9%	4.6%	0.4%	70.4%	70%	5392
Overall Correctly Classified														**69%**	**103393**

Table C.2: The confusion matrix based on the **Random Tree** classifier without feature selection, using **the combined features** on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	52.4%	9.6%	7.6%	1.0%	1.9%	4.9%	1.5%	4.5%	1.7%	5.0%	2.5%	5.7%	1.7%	52%	7845
b=Nanday Parakeet	13.1%	49.5%	6.5%	2.1%	3.4%	5.0%	0.9%	3.3%	4.1%	3.0%	3.0%	3.5%	2.5%	50%	6246
c=Blue-crowned Parakeet	10.7%	8.1%	59.0%	1.3%	2.7%	3.7%	0.9%	4.0%	2.3%	1.8%	1.4%	2.3%	1.7%	59%	5332
d=Common House Martin	0.4%	0.7%	0.3%	83.7%	0.2%	0.4%	4.1%	0.5%	0.5%	2.1%	2.1%	0.9%	4.0%	84%	17896
e=Eastern Rosella	5.2%	8.2%	3.8%	1.3%	43.9%	2.7%	1.7%	4.0%	1.9%	10.1%	5.2%	10.2%	2.0%	44%	3247
f=Budgerigar (yellow)	9.0%	7.0%	4.1%	1.8%	2.0%	51.1%	3.3%	6.1%	1.4%	4.6%	3.3%	3.4%	2.9%	51%	4329
g=House Sparrow	2.3%	1.1%	0.9%	14.0%	1.1%	2.9%	43.6%	2.5%	0.9%	6.8%	4.5%	3.2%	16.3%	44%	5318
h=Budgerigar (wild-type)	10.7%	6.8%	7.0%	2.4%	3.4%	7.1%	3.7%	33.5%	2.9%	8.7%	4.5%	6.4%	2.8%	33%	3349
i=Common Wood Pigeon	7.0%	13.7%	6.0%	3.6%	2.9%	2.8%	2.2%	4.1%	38.6%	4.7%	5.7%	3.7%	5.2%	39%	2027
j=Black-headed Gull	1.2%	0.6%	0.3%	1.2%	1.1%	0.8%	1.1%	1.0%	0.3%	85.4%	1.9%	4.6%	0.5%	85%	29695
k=Cockatiel	3.6%	3.1%	1.1%	7.5%	3.1%	2.7%	3.8%	2.7%	1.8%	9.0%	49.1%	6.1%	6.6%	49%	5687
l=Budgerigar (blue)	6.0%	3.4%	1.8%	2.1%	4.1%	2.4%	2.6%	2.8%	1.2%	18.2%	4.6%	49.3%	1.6%	49%	7030
m=Common Starling	2.1%	2.2%	1.7%	12.9%	1.0%	2.5%	16.2%	1.7%	2.0%	2.5%	7.9%	2.4%	44.9%	45%	5392
Overall Correctly Classified														**65%**	**103393**

Table C.3: The confusion matrix based on the **SVM** classifier without feature selection, using **the combined features** on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	59.8%	11.0%	6.8%	0.5%	0.6%	6.6%	0.9%	0.9%	0.7%	2.3%	2.6%	6.5%	0.7%	60%	7845
b=Nanday Parakeet	14.5%	56.3%	8.4%	0.8%	1.3%	4.2%	0.4%	0.6%	3.5%	1.0%	4.1%	3.4%	1.4%	56%	6246
c=Blue-crowned Parakeet	15.5%	7.1%	66.0%	0.8%	1.3%	3.2%	0.4%	0.7%	0.5%	0.5%	0.3%	3.1%	0.8%	66%	5332
d=Common House Martin	0.3%	0.4%	0.1%	97.5%	0.0%	0.0%	0.4%	0.0%	0.1%	0.2%	0.3%	0.1%	0.6%	98%	17896
e=Eastern Rosella	4.3%	7.5%	6.3%	1.2%	43.6%	2.8%	1.3%	0.6%	0.7%	12.9%	4.0%	13.7%	1.2%	44%	3247
f=Budgerigar (yellow)	5.2%	1.8%	2.4%	1.9%	1.0%	78.1%	1.2%	1.4%	0.2%	3.6%	1.9%	0.9%	0.4%	78%	4329
g=House Sparrow	0.9%	0.1%	0.5%	15.0%	0.9%	2.3%	51.6%	0.5%	0.3%	6.5%	2.4%	1.9%	17.0%	52%	5318
h=Budgerigar (wild-type)	11.7%	4.1%	8.2%	2.6%	3.4%	3.6%	4.6%	33.2%	2.7%	11.3%	7.7%	6.4%	0.6%	33%	3349
i=Common Wood Pigeon	6.6%	14.2%	5.6%	4.7%	2.6%	1.1%	0.4%	0.2%	48.1%	0.8%	8.4%	1.9%	5.3%	48%	2027
j=Black-headed Gull	0.2%	0.1%	0.0%	0.3%	0.1%	0.0%	0.3%	0.0%	0.0%	96.8%	0.5%	1.5%	0.0%	97%	29695
k=Cockatiel	2.1%	1.4%	0.3%	6.2%	1.0%	1.3%	0.8%	0.3%	0.8%	4.7%	68.2%	5.0%	7.7%	68%	5687
l=Budgerigar (blue)	5.8%	2.2%	1.4%	2.3%	2.3%	1.9%	2.0%	0.7%	0.3%	22.5%	5.0%	52.8%	0.8%	53%	7030
m=Common Starling	1.4%	0.6%	0.5%	8.1%	0.3%	0.9%	7.8%	0.1%	0.5%	1.0%	3.8%	0.4%	74.6%	75%	5392
Overall Correctly Classified														**77%**	**103393**

Appendix D

Appearance Set with Feature Selection

Table D.1: The confusion matrix based on the **Random Tree** classifier using **the selected features from the appearance set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	66.0%	8.0%	5.8%	0.4%	1.2%	3.4%	1.0%	4.0%	1.4%	2.1%	1.7%	4.4%	0.7%	66%	12801
b=Nanday Parakeet	11.6%	61.7%	5.2%	0.9%	2.1%	3.4%	0.6%	3.3%	5.1%	0.8%	1.6%	2.3%	1.6%	62%	10025
c=Blue-crowned Parakeet	8.2%	5.9%	70.5%	0.5%	2.0%	2.5%	1.1%	3.4%	1.8%	0.4%	0.7%	1.9%	1.0%	71%	9076
d=Common House Martin	0.2%	0.5%	0.3%	87.9%	0.1%	0.2%	3.9%	0.4%	0.5%	0.9%	1.5%	0.4%	3.1%	88%	25517
e=Eastern Rosella	3.2%	3.6%	3.9%	0.6%	61.6%	1.5%	1.9%	3.2%	1.0%	4.3%	4.0%	9.7%	1.3%	62%	5929
f=Budgerigar (yellow)	6.4%	4.3%	3.5%	0.5%	1.2%	66.6%	2.5%	5.8%	1.2%	2.0%	1.8%	2.2%	2.0%	67%	7667
g=House Sparrow	1.1%	0.4%	0.8%	9.7%	1.1%	1.9%	56.0%	2.2%	0.6%	5.3%	3.1%	3.0%	14.8%	56%	10191
h=Budgerigar (wild-type)	9.4%	5.3%	5.6%	1.7%	3.4%	6.4%	4.0%	45.7%	3.0%	3.7%	3.5%	5.8%	2.7%	46%	6283
i=Common Wood Pigeon	4.3%	10.6%	4.0%	2.3%	2.2%	2.0%	1.2%	3.5%	59.5%	1.3%	2.5%	2.3%	4.3%	60%	4301
j=Black-headed Gull	0.7%	0.3%	0.1%	0.5%	0.7%	0.5%	1.4%	0.5%	0.2%	90.6%	0.7%	3.6%	0.2%	91%	38764
k=Cockatiel	2.6%	1.8%	0.8%	4.1%	2.3%	1.4%	2.7%	2.4%	1.4%	3.0%	66.9%	4.2%	6.4%	67%	9398
l=Budgerigar (blue)	4.6%	1.7%	1.6%	1.1%	4.4%	1.3%	2.7%	3.3%	0.8%	11.8%	3.1%	62.4%	1.3%	62%	12090
m=Common Starling	0.9%	1.4%	1.1%	8.3%	0.9%	1.3%	14.7%	1.7%	2.1%	1.0%	6.4%	1.7%	58.4%	58%	9865
Overall Correctly Classified														**73%**	**161907**

Table D.2: The confusion matrix based on the **Naive Bayes** classifier using **the selected features from the appearance set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	29.1%	17.5%	8.5%	0.5%	1.7%	15.9%	1.3%	6.5%	2.9%	3.2%	2.9%	9.2%	0.8%	29%	12801
b=Nanday Parakeet	6.5%	46.2%	6.9%	0.9%	4.4%	6.4%	0.2%	1.1%	12.7%	2.0%	6.0%	3.9%	2.7%	46%	10025
c=Blue-crowned Parakeet	9.9%	11.0%	52.6%	1.0%	2.4%	8.7%	0.8%	5.7%	2.4%	0.5%	1.2%	2.5%	1.4%	53%	9076
d=Common House Martin	0.0%	0.7%	0.3%	85.8%	0.0%	0.0%	4.8%	0.1%	0.2%	0.0%	0.7%	0.1%	7.3%	86%	25517
e=Eastern Rosella	0.7%	9.4%	3.8%	1.5%	53.1%	4.7%	1.4%	2.3%	1.9%	3.6%	4.3%	12.2%	1.2%	53%	5929
f=Budgerigar (yellow)	2.8%	4.1%	4.2%	0.3%	0.7%	63.6%	1.8%	7.2%	3.0%	2.3%	7.3%	0.5%	2.2%	64%	7667
g=House Sparrow	0.6%	0.1%	0.7%	16.6%	0.3%	2.1%	41.6%	2.3%	0.0%	5.1%	4.6%	4.9%	21.0%	42%	10191
h=Budgerigar (wild-type)	6.2%	6.6%	9.0%	2.9%	3.3%	14.7%	4.8%	19.5%	5.9%	6.2%	11.0%	7.2%	2.8%	19%	6283
i=Common Wood Pigeon	3.6%	20.6%	4.1%	4.5%	1.7%	1.6%	1.1%	1.5%	45.8%	0.8%	6.1%	1.8%	6.8%	46%	4301
j=Black-headed Gull	2.7%	1.6%	0.2%	0.7%	1.2%	0.3%	4.3%	6.8%	0.9%	70.0%	3.8%	7.5%	0.2%	70%	38764
k=Cockatiel	1.0%	3.3%	0.7%	8.8%	2.7%	2.6%	3.0%	3.0%	1.2%	2.7%	45.4%	11.8%	13.8%	45%	9398
l=Budgerigar (blue)	1.9%	5.7%	0.8%	3.3%	11.1%	0.7%	2.3%	4.7%	1.8%	23.3%	6.2%	37.6%	0.3%	38%	12090
m=Common Starling	0.8%	0.8%	1.0%	17.4%	0.0%	3.5%	15.1%	0.9%	0.7%	0.2%	5.5%	0.1%	53.8%	54%	9865
Overall Correctly Classified														**57%**	**161907**

Table D.3: The confusion matrix based on the **SVM** classifier using **the selected features from the appearance set** based on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	53.0%	8.9%	8.0%	0.7%	0.3%	10.7%	0.5%	0.3%	0.7%	5.0%	2.0%	8.5%	1.3%	53%	12801
b=Nanday Parakeet	18.8%	39.4%	12.2%	2.0%	1.8%	6.6%	0.3%	0.2%	5.7%	3.4%	4.8%	3.6%	1.4%	39%	10025
c=Blue-crowned Parakeet	21.3%	10.5%	49.3%	1.2%	1.4%	6.1%	0.5%	0.2%	0.5%	3.3%	1.2%	2.8%	1.7%	49%	9076
d=Common House Martin	0.3%	0.4%	0.1%	95.7%	0.0%	0.1%	0.9%	0.0%	0.0%	0.7%	0.5%	0.1%	1.2%	96%	25517
e=Eastern Rosella	4.4%	8.5%	5.5%	1.3%	38.8%	2.6%	0.9%	0.1%	0.5%	14.0%	6.7%	14.5%	2.3%	39%	5929
f=Budgerigar (yellow)	7.9%	2.8%	2.9%	1.3%	0.9%	67.1%	1.0%	0.2%	0.5%	8.8%	4.4%	0.3%	1.8%	67%	7667
g=House Sparrow	0.4%	0.2%	0.2%	21.1%	0.9%	2.8%	34.7%	0.1%	0.2%	14.0%	3.1%	1.4%	20.8%	35%	10191
h=Budgerigar (wild-type)	17.4%	5.0%	8.6%	3.4%	4.1%	12.0%	5.4%	14.4%	3.1%	10.2%	9.7%	3.6%	3.1%	14%	6283
i=Common Wood Pigeon	10.0%	22.2%	6.9%	5.4%	2.7%	0.8%	0.4%	0.4%	31.8%	3.8%	5.0%	2.5%	8.0%	32%	4301
j=Black-headed Gull	0.6%	0.3%	0.2%	0.5%	0.0%	0.2%	1.1%	0.1%	0.1%	94.1%	0.5%	2.4%	0.0%	94%	38764
k=Cockatiel	2.9%	1.0%	0.7%	8.6%	2.1%	1.7%	1.2%	0.1%	0.6%	4.4%	51.8%	6.6%	18.3%	52%	9398
l=Budgerigar (blue)	6.4%	3.1%	0.9%	1.8%	4.5%	0.7%	1.9%	0.1%	0.4%	32.9%	7.5%	38.5%	1.2%	39%	12090
m=Common Starling	1.5%	0.5%	0.7%	12.9%	0.8%	1.4%	5.6%	0.0%	0.2%	1.2%	2.2%	0.8%	72.2%	72%	9865
Overall Correctly Classified														**65%**	**161907**

Appendix E

Motion Set with Feature Selection

Table E.1: The confusion matrix based on the **Random Tree** classifier using **the selected features from the motion set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	16.4%	9.7%	7.8%	13.4%	3.7%	4.0%	5.0%	3.4%	2.8%	14.2%	7.6%	6.3%	5.8%	16%	7845
b=Nanday Parakeet	10.8%	13.8%	8.7%	12.7%	4.5%	4.4%	5.6%	3.3%	3.4%	12.4%	8.0%	5.6%	6.8%	14%	6246
c=Blue-crowned Parakeet	12.1%	10.5%	11.1%	13.9%	4.1%	3.8%	5.2%	3.6%	3.1%	11.0%	8.5%	5.4%	7.8%	11%	5332
d=Common House Martin	5.7%	4.4%	3.9%	35.1%	2.7%	4.2%	4.0%	2.9%	1.3%	20.9%	4.5%	6.7%	3.8%	35%	17896
e=Eastern Rosella	10.2%	8.3%	5.6%	15.0%	8.3%	4.2%	4.9%	3.3%	2.6%	18.3%	6.6%	7.0%	5.6%	8%	3247
f=Budgerigar (yellow)	6.6%	5.5%	3.7%	15.8%	3.0%	11.9%	5.3%	4.8%	1.9%	18.7%	5.8%	12.1%	4.9%	12%	4329
g=House Sparrow	7.3%	5.9%	5.2%	14.5%	3.1%	4.4%	17.7%	4.0%	2.2%	16.9%	4.6%	7.8%	6.3%	18%	5318
h=Budgerigar (wild-type)	6.8%	6.2%	4.6%	15.3%	2.8%	6.5%	6.7%	10.7%	2.0%	18.5%	5.0%	9.7%	5.2%	11%	3349
i=Common Wood Pigeon	11.1%	10.1%	9.3%	11.7%	4.7%	4.6%	3.9%	3.7%	7.3%	13.4%	7.7%	5.3%	7.2%	7%	2027
j=Black-headed Gull	3.7%	2.6%	1.9%	12.9%	2.0%	3.1%	2.8%	2.2%	1.0%	57.8%	3.6%	4.7%	1.7%	58%	29695
k=Cockatiel	9.7%	8.8%	7.4%	13.6%	3.5%	3.8%	4.4%	3.7%	2.6%	18.4%	12.1%	6.0%	6.1%	12%	5687
l=Budgerigar (blue)	6.3%	4.6%	3.9%	16.7%	3.0%	6.9%	5.3%	5.2%	1.5%	19.6%	4.7%	18.8%	3.6%	19%	7030
m=Common Starling	9.6%	9.1%	6.9%	12.9%	3.7%	3.8%	6.6%	3.0%	2.8%	9.4%	6.0%	5.7%	20.3%	20%	5392
Overall Correctly Classified														**30%**	**103393**

Table E.2: The confusion matrix based on the **SVM** classifier using **the selected features from the motion set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	12.8%	3.1%	0.0%	32.2%	0.0%	0.0%	0.1%	0.0%	0.0%	39.0%	0.1%	0.1%	12.6%	13%	7845
b=Nanday Parakeet	12.5%	15.8%	0.0%	27.5%	0.0%	0.0%	0.1%	0.0%	0.0%	35.8%	0.0%	0.0%	8.3%	16%	6246
c=Blue-crowned Parakeet	13.2%	4.9%	16.9%	22.2%	0.0%	0.0%	0.1%	0.0%	0.0%	32.1%	0.1%	0.1%	10.4%	17%	5332
d=Common House Martin	1.8%	0.5%	0.0%	25.3%	0.0%	0.0%	0.0%	0.0%	0.0%	70.7%	0.0%	0.0%	1.6%	25%	17896
e=Eastern Rosella	6.8%	1.8%	0.0%	32.9%	15.4%	0.0%	0.0%	0.0%	0.0%	36.7%	0.1%	0.0%	6.3%	15%	3247
f=Budgerigar (yellow)	2.5%	0.8%	0.0%	29.2%	0.0%	16.5%	0.0%	0.0%	0.0%	46.8%	0.0%	0.0%	4.1%	16%	4329
g=House Sparrow	8.4%	2.1%	0.0%	32.5%	0.0%	0.0%	11.9%	0.0%	0.0%	31.9%	0.0%	0.0%	13.1%	12%	5318
h=Budgerigar (wild-type)	4.7%	1.7%	0.0%	29.0%	0.0%	0.0%	0.1%	15.1%	0.0%	45.8%	0.1%	0.0%	3.5%	15%	3349
i=Common Wood Pigeon	18.5%	5.4%	0.0%	24.2%	0.0%	0.0%	0.1%	0.0%	20.2%	22.3%	0.1%	0.2%	8.8%	20%	2027
j=Black-headed Gull	0.5%	0.3%	0.0%	30.0%	0.0%	0.0%	0.0%	0.0%	0.0%	68.6%	0.0%	0.0%	0.5%	69%	29695
k=Cockatiel	5.9%	3.1%	0.0%	24.8%	0.0%	0.0%	0.0%	0.0%	0.0%	41.5%	21.2%	0.0%	3.4%	21%	5687
l=Budgerigar (blue)	3.0%	0.8%	0.0%	26.2%	0.0%	0.0%	0.0%	0.0%	0.0%	53.0%	0.0%	14.2%	2.8%	14%	7030
m=Common Starling	10.1%	4.1%	0.0%	34.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.6%	0.1%	0.1%	28.0%	28%	5392
Overall Correctly Classified														**33%**	**103393**

Table E.3: The confusion matrix based on the **Naive Bayes** classifier using **the selected features from the motion set** based on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	19.7%	12.4%	8.5%	14.0%	1.2%	1.0%	5.5%	0.4%	2.2%	16.4%	9.7%	2.4%	6.8%	20%	7845
b=Nanday Parakeet	12.2%	24.0%	7.7%	10.5%	0.8%	0.9%	5.4%	0.6%	2.1%	13.4%	10.5%	2.2%	9.6%	24%	6246
c=Blue-crowned Parakeet	14.3%	13.6%	17.1%	11.6%	1.5%	1.0%	5.5%	0.9%	1.4%	9.6%	10.4%	1.8%	11.4%	17%	5332
d=Common House Martin	3.7%	3.1%	2.6%	37.3%	0.4%	2.2%	2.7%	1.1%	0.9%	33.8%	2.2%	5.3%	4.8%	37%	17896
e=Eastern Rosella	9.4%	7.8%	6.6%	21.4%	5.9%	1.1%	3.4%	0.6%	0.7%	24.6%	7.3%	2.4%	8.8%	6%	3247
f=Budgerigar (yellow)	2.7%	1.3%	2.3%	11.9%	0.8%	13.1%	4.7%	2.1%	0.9%	25.2%	2.1%	28.3%	4.6%	13%	4329
g=House Sparrow	4.1%	2.9%	1.9%	8.1%	0.4%	1.9%	45.1%	1.4%	0.5%	18.1%	3.3%	6.0%	6.4%	45%	5318
h=Budgerigar (wild-type)	3.1%	3.6%	2.7%	13.9%	0.8%	5.6%	6.5%	8.6%	0.8%	23.8%	3.1%	18.2%	9.5%	9%	3349
i=Common Wood Pigeon	17.7%	12.7%	7.7%	11.1%	1.0%	2.4%	3.5%	1.1%	11.4%	11.2%	8.2%	3.1%	8.9%	11%	2027
j=Black-headed Gull	2.4%	0.8%	0.5%	18.9%	0.2%	1.1%	1.9%	0.2%	0.3%	68.5%	2.0%	2.1%	1.0%	69%	29695
k=Cockatiel	9.0%	8.5%	4.7%	13.1%	0.6%	1.7%	4.5%	0.7%	1.1%	25.4%	21.2%	3.2%	6.3%	21%	5687
l=Budgerigar (blue)	2.9%	1.5%	1.2%	16.4%	0.5%	6.7%	4.1%	2.7%	0.5%	23.8%	2.3%	32.3%	5.1%	32%	7030
m=Common Starling	4.2%	6.3%	2.9%	15.7%	0.6%	1.1%	4.0%	1.0%	1.1%	3.6%	3.9%	2.4%	53.1%	53%	5392
Overall Correctly Classified														**40%**	**103393**

Appendix F

Combined Set with Feature Selection

Table F.1: The confusion matrix based on the **Random Tree** classifier using **the selected features from the combined set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	64.0%	8.8%	5.9%	0.5%	1.6%	3.6%	0.9%	4.0%	1.3%	2.6%	1.9%	4.0%	0.9%	64%	7845
b=Nanday Parakeet	11.9%	61.2%	5.4%	1.4%	2.1%	3.4%	0.5%	2.8%	4.1%	1.6%	1.9%	2.1%	1.5%	61%	6246
c=Blue-crowned Parakeet	8.7%	6.3%	69.3%	0.6%	1.8%	2.1%	0.8%	3.9%	2.4%	0.5%	1.0%	1.8%	0.8%	69%	5332
d=Common House Martin	0.2%	0.5%	0.2%	88.6%	0.2%	0.2%	3.1%	0.3%	0.5%	1.0%	1.7%	0.6%	3.0%	89%	17896
e=Eastern Rosella	4.1%	3.8%	2.8%	1.0%	58.9%	2.3%	2.0%	2.4%	1.0%	6.5%	4.6%	9.2%	1.4%	59%	3247
f=Budgerigar (yellow)	8.0%	5.6%	3.1%	0.7%	1.5%	62.7%	2.6%	4.7%	1.1%	2.7%	2.5%	2.9%	2.1%	63%	4329
g=House Sparrow	1.6%	0.6%	0.7%	11.7%	1.0%	2.1%	53.4%	2.2%	0.4%	6.0%	3.0%	2.5%	14.6%	53%	5318
h=Budgerigar (wild-type)	10.5%	6.3%	6.1%	1.4%	3.1%	6.8%	2.6%	44.2%	3.2%	5.0%	3.4%	5.6%	1.8%	44%	3349
i=Common Wood Pigeon	5.4%	11.8%	5.5%	3.6%	2.2%	2.2%	1.6%	3.9%	51.1%	1.9%	3.6%	2.6%	4.4%	51%	2027
j=Black-headed Gull	0.7%	0.3%	0.2%	0.6%	0.6%	0.3%	0.9%	0.5%	0.1%	91.5%	0.9%	3.1%	0.2%	92%	29695
k=Cockatiel	2.6%	2.1%	0.9%	5.5%	2.4%	1.7%	2.7%	2.1%	1.2%	4.5%	64.5%	4.3%	5.5%	65%	5687
l=Budgerigar (blue)	4.4%	2.1%	1.1%	1.2%	3.8%	1.4%	1.8%	2.0%	0.8%	13.4%	3.7%	63.1%	1.3%	63%	7030
m=Common Starling	1.4%	1.5%	1.0%	9.8%	0.8%	1.6%	15.0%	1.1%	1.7%	1.5%	6.5%	1.4%	56.6%	57%	5392
Overall Correctly Classified														**74%**	**103393**

Table F.2: The confusion matrix based on the **SVM** classifier using **the selected features from the combined set** based on the **thirteen classes dataset**

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Samples
a=Alexandrine Parakeet	55.6%	9.8%	8.9%	0.7%	0.2%	8.5%	0.7%	0.3%	0.4%	3.8%	2.5%	7.7%	0.8%	56%	7845
b=Nanday Parakeet	19.0%	43.9%	12.1%	2.0%	1.0%	5.4%	0.4%	0.3%	3.6%	2.7%	4.7%	3.2%	1.7%	44%	6246
c=Blue-crowned Parakeet	22.1%	11.9%	51.9%	0.7%	0.6%	4.7%	0.4%	0.4%	0.3%	1.9%	0.5%	3.1%	1.5%	52%	5332
d=Common House Martin	0.2%	0.7%	0.0%	97.4%	0.0%	0.1%	0.3%	0.0%	0.0%	0.4%	0.3%	0.0%	0.6%	97%	17896
e=Eastern Rosella	3.3%	10.7%	4.6%	1.8%	34.2%	3.9%	1.2%	0.1%	0.3%	16.9%	7.9%	13.6%	1.6%	34%	3247
f=Budgerigar (yellow)	6.3%	1.9%	2.0%	2.4%	0.9%	73.2%	1.1%	0.6%	0.2%	7.0%	3.0%	0.5%	0.9%	73%	4329
g=House Sparrow	0.9%	0.4%	0.2%	20.5%	0.3%	2.7%	39.9%	0.2%	0.2%	9.7%	2.7%	1.1%	21.3%	40%	5318
h=Budgerigar (wild-type)	12.8%	5.0%	5.9%	3.4%	3.4%	6.0%	3.2%	26.2%	2.0%	17.9%	9.4%	3.9%	0.9%	26%	3349
i=Common Wood Pigeon	9.2%	29.4%	7.0%	5.7%	1.7%	1.1%	0.5%	0.5%	21.7%	3.5%	11.4%	2.3%	6.1%	22%	2027
j=Black-headed Gull	0.3%	0.3%	0.1%	0.6%	0.0%	0.1%	0.5%	0.0%	0.0%	96.3%	0.4%	1.5%	0.0%	96%	29695
k=Cockatiel	3.0%	1.3%	0.6%	9.3%	1.3%	2.4%	1.3%	0.2%	0.7%	5.6%	58.0%	4.9%	11.4%	58%	5687
l=Budgerigar (blue)	6.1%	2.7%	1.0%	2.4%	2.9%	1.4%	1.5%	0.3%	0.1%	32.2%	7.9%	40.9%	0.6%	41%	7030
m=Common Starling	1.7%	1.1%	0.8%	11.7%	0.3%	1.3%	7.3%	0.1%	0.2%	0.6%	2.7%	0.7%	71.6%	72%	5392
Overall Correctly Classified														**71%**	**103393**

Table F.3: The confusion matrix based on the **Naive Bayes** classifier using **the selected features from the combined set** based on the **thirteen classes dataset**

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Samples**
a=Alexandrine Parakeet	39.3%	18.4%	8.8%	0.5%	1.7%	7.9%	1.0%	6.5%	2.7%	1.8%	4.2%	6.7%	0.5%	39%	7845
b=Nanday Parakeet	9.5%	53.3%	5.7%	0.8%	3.7%	3.5%	0.5%	0.9%	10.7%	1.0%	5.0%	3.3%	2.2%	53%	6246
c=Blue-crowned Parakeet	13.3%	10.5%	57.6%	0.6%	2.1%	4.5%	0.4%	4.8%	2.6%	0.3%	1.2%	1.3%	0.8%	58%	5332
d=Common House Martin	0.0%	0.8%	0.2%	90.6%	0.0%	0.0%	4.6%	0.0%	0.2%	0.0%	0.5%	0.1%	3.0%	91%	17896
e=Eastern Rosella	1.5%	7.6%	3.1%	1.9%	56.8%	5.7%	1.2%	1.6%	1.5%	4.7%	4.9%	8.9%	0.6%	57%	3247
f=Budgerigar (yellow)	1.9%	3.1%	2.1%	0.4%	0.4%	72.4%	1.6%	8.0%	2.1%	0.4%	5.0%	1.6%	1.0%	72%	4329
g=House Sparrow	0.5%	0.1%	0.7%	15.0%	0.3%	1.6%	57.2%	1.8%	0.0%	4.4%	2.6%	3.4%	12.4%	57%	5318
h=Budgerigar (wild-type)	5.0%	5.6%	5.6%	2.5%	1.8%	15.7%	2.7%	30.6%	5.0%	3.3%	11.1%	9.8%	1.4%	31%	3349
i=Common Wood Pigeon	3.2%	17.0%	4.1%	4.6%	0.8%	1.2%	0.6%	1.5%	49.5%	0.3%	9.1%	1.7%	6.3%	50%	2027
j=Black-headed Gull	1.5%	1.3%	0.1%	0.9%	0.9%	0.2%	2.6%	2.0%	0.3%	81.6%	3.9%	4.7%	0.0%	82%	29695
k=Cockatiel	2.0%	3.1%	0.7%	8.4%	3.3%	2.1%	2.8%	1.8%	0.7%	3.1%	54.5%	7.1%	10.4%	54%	5687
l=Budgerigar (blue)	1.7%	3.6%	0.8%	2.6%	4.3%	1.2%	1.4%	5.8%	1.7%	13.0%	7.0%	56.7%	0.3%	57%	7030
m=Common Starling	1.4%	0.7%	0.4%	12.4%	0.0%	1.9%	9.8%	0.4%	0.5%	0.8%	5.3%	0.2%	66.1%	66%	5392
Overall Correctly Classified														**68%**	**103393**

Appendix G

Majority Voting Results

Table G.1: Confusion matrix of video classification using our Inflight #2 Dataset with Naive Bayes Classifier

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Videos
a=Alexandrine Parakeet	96.1%	1.3%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	0.0%	0.0%	96%	77
b=Nanday Parakeet	1.7%	96.6%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	59
c=Blue-crowned Parakeet	1.7%	0.0%	96.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.0%	0.0%	97%	58
d=Common House Martin	0.0%	0.0%	0.0%	99.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	0.0%	0.0%	0.0%	99%	114
e=Eastern Rosella	0.0%	0.0%	2.5%	0.0%	92.5%	0.0%	0.0%	5.0%	0.0%	0.0%	0.0%	0.0%	0.0%	93%	40
f=Budgerigar (yellow)	2.1%	0.0%	2.1%	0.0%	0.0%	95.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	96%	47
g=House Sparrow	0.0%	0.0%	0.0%	2.7%	0.0%	0.0%	97.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	74
h=Budgerigar (wild-type)	0.0%	0.0%	0.0%	0.0%	0.0%	12.2%	0.0%	85.4%	0.0%	0.0%	0.0%	2.4%	0.0%	85%	41
i=Common Wood Pigeon	0.0%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	0.0%	88.6%	2.9%	0.0%	0.0%	2.9%	89%	35
j=Black-headed Gull	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100%	142
k=Cockatiel	0.0%	0.0%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98.3%	0.0%	0.0%	98%	58
l=Budgerigar (blue)	0.0%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	2.6%	0.0%	94.7%	0.0%	95%	76
m=Common Starling	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	94.4%	94%	71
Overall Correctly Classified														95%	892

Table G.2: Confusion matrix of video classification using our Inflight #2 Dataset with SVM Classifier.

	a	b	c	d	e	f	g	h	i	j	k	l	m	%CC	Videos
a=Alexandrine Parakeet	98.7%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	99%	77
b=Nanday Parakeet	6.8%	93.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	93%	59
c=Blue-crowned Parakeet	1.7%	0.0%	98.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98%	58
d=Common House Martin	0.0%	0.9%	0.0%	97.4%	0.0%	0.0%	0.9%	0.0%	0.0%	0.9%	0.0%	0.0%	0.0%	97%	114
e=Eastern Rosella	0.0%	0.0%	2.5%	0.0%	90.0%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	5.0%	0.0%	90%	40
f=Budgerigar (yellow)	0.0%	0.0%	0.0%	0.0%	0.0%	95.7%	2.1%	0.0%	2.1%	0.0%	0.0%	0.0%	0.0%	96%	47
g=House Sparrow	0.0%	0.0%	0.0%	8.1%	0.0%	0.0%	89.2%	0.0%	0.0%	0.0%	0.0%	0.0%	2.7%	89%	74
h=Budgerigar (wild-type)	14.6%	0.0%	9.8%	0.0%	0.0%	7.3%	0.0%	58.5%	0.0%	9.8%	0.0%	0.0%	0.0%	59%	41
i=Common Wood Pigeon	0.0%	40.0%	2.9%	0.0%	0.0%	0.0%	0.0%	0.0%	57.1%	0.0%	0.0%	0.0%	0.0%	57%	35
j=Black-headed Gull	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100%	142
k=Cockatiel	0.0%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98.3%	0.0%	0.0%	98%	58
l=Budgerigar (blue)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	21.1%	0.0%	78.9%	0.0%	79%	76
m=Common Starling	0.0%	0.0%	0.0%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	0.0%	0.0%	97.2%	97%	71
Overall Correctly Classified														**89%**	**892**

Table G.3: Confusion matrix of video classification using our Inflight #2 Dataset with Random Tree Classifier

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**%CC**	**Videos**
a=Alexandrine Parakeet	97.4%	2.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	77
b=Nanday Parakeet	0.0%	98.3%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98%	59
c=Blue-crowned Parakeet	1.7%	1.7%	96.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97%	58
d=Common House Martin	0.0%	0.0%	0.0%	97.4%	0.0%	0.0%	0.9%	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%	97%	114
e=Eastern Rosella	0.0%	0.0%	0.0%	2.5%	95.0%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	0.0%	0.0%	95%	40
f=Budgerigar (yellow)	0.0%	0.0%	0.0%	0.0%	0.0%	95.7%	0.0%	2.1%	0.0%	2.1%	0.0%	0.0%	0.0%	96%	47
g=House Sparrow	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	98.6%	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	99%	74
h=Budgerigar (wild-type)	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	97.6%	0.0%	0.0%	0.0%	0.0%	0.0%	98%	41
i=Common Wood Pigeon	0.0%	0.0%	0.0%	2.9%	0.0%	0.0%	0.0%	0.0%	88.6%	5.7%	0.0%	0.0%	2.9%	89%	35
j=Black-headed Gull	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100%	142
k=Cockatiel	0.0%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	96.6%	0.0%	1.7%	97%	58
l=Budgerigar (blue)	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	1.3%	0.0%	97.4%	0.0%	97%	76
m=Common Starling	1.4%	0.0%	0.0%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	94.4%	94%	71
Overall Correctly Classified														**96%**	**892**

Appendix H

Significance Test Results

Table H.1: Wilcoxon's sign rank test based on 10 times 5-fold cross validation. This result is based on dataset #2 using the appearance feature set without feature selection.

				DIFFERENCE			SIGN			ABS			RANK			SiGN X RANK		
NB	**RF**	**RT**	**SVM**	**RF-NB**	**RF-RT**	**RF-SVM**	**RF-NB**	**RF-RT**	**RF-SVM**	**|RF-NB|**	**|RF-RT|**	**|RF-SVM|**	**RF-NB**	**RF-RT**	**RF-SVM**	**RF-NB**	**RF-RT**	**RF-SVM**
54.01	86.64	69.44	69.69	32.64	17.20	16.95	1	1	1	33	17	17	44	13	37.5	44	13	37.5
53.91	86.61	67.71	69.76	32.69	18.89	16.85	1	1	1	33	19	17	44	44.5	37	44	44.5	37
53.86	86.15	67.67	69.81	32.29	18.48	16.35	1	1	1	32	18	16	19.5	29.5	12.5	19.5	29.5	12.5
53.79	85.92	68.61	69.55	32.13	17.31	16.37	1	1	1	32	17	16	19	12.5	12	19	12.5	12
53.86	86.20	67.95	69.67	32.34	18.25	16.53	1	1	1	32	18	17	18.5	28	34.5	18.5	28	34.5
54.10	86.27	66.73	69.82	32.17	19.54	16.45	1	1	1	32	20	16	18	45	11.5	18	45	11.5
53.51	86.03	67.88	69.81	32.52	18.15	16.22	1	1	1	33	18	16	39.5	27.5	11	39.5	27.5	11
53.43	86.33	67.55	69.47	32.90	18.78	16.86	1	1	1	33	19	17	39	40	32	39	40	32
54.27	86.48	70.98	69.81	32.21	15.50	16.67	1	1	1	32	15	17	17.5	2.5	31.5	17.5	2.5	31.5
54.08	86.23	67.65	69.51	32.15	18.58	16.73	1	1	1	32	19	17	17	38.5	31	17	38.5	31
53.66	86.21	67.11	69.62	32.55	19.10	16.59	1	1	1	33	19	17	36.5	38	30.5	36.5	38	30.5
53.89	86.27	68.45	70.00	32.38	17.81	16.27	1	1	1	32	18	16	16.5	26	10.5	16.5	26	10.5
53.90	86.12	68.38	69.34	32.22	17.74	16.78	1	1	1	32	18	17	16	25.5	29	16	25.5	29
53.78	86.06	68.88	69.53	32.28	17.18	16.53	1	1	1	32	17	17	15.5	11	28.5	15.5	11	28.5
54.21	86.29	69.11	69.94	32.08	17.17	16.35	1	1	1	32	17	16	15	10.5	10	15	10.5	10
53.72	86.18	67.97	69.90	32.46	18.21	16.29	1	1	1	32	18	16	14.5	23	9.5	14.5	23	9.5
54.01	86.23	71.15	69.69	32.23	15.08	16.54	1	1	1	32	15	17	14	2	26	14	2	26
53.67	86.14	67.18	69.56	32.47	18.96	16.58	1	1	1	32	19	17	13.5	31.5	25.5	13.5	31.5	25.5
53.68	85.98	68.79	69.52	32.30	17.18	16.46	1	1	1	32	17	16	13	9	9	13	9	9
54.22	86.31	69.35	69.79	32.10	16.96	16.52	1	1	1	32	17	17	12.5	8.5	24	12.5	8.5	24
54.21	85.95	68.20	69.66	31.74	17.74	16.29	1	1	1	32	18	16	12	19.5	8.5	12	19.5	8.5
53.41	85.88	69.00	69.63	32.47	16.87	16.24	1	1	1	32	17	16	11.5	8	8	11.5	8	8
53.64	85.80	69.19	69.63	32.16	16.62	16.17	1	1	1	32	17	16	11	7.5	7.5	11	7.5	7.5
53.70	86.31	68.56	69.60	32.61	17.76	16.71	1	1	1	33	18	17	24	17	20.5	24	17	20.5
54.36	86.43	68.87	69.94	32.07	17.55	16.49	1	1	1	32	18	16	10.5	16.5	7	10.5	16.5	7
53.68	86.40	69.64	69.53	32.72	16.76	16.87	1	1	1	33	17	17	22.5	7	19	22.5	7	19
54.19	86.16	69.58	69.68	31.97	16.58	16.48	1	1	1	32	17	16	10	6.5	6.5	10	6.5	6.5
53.69	86.09	69.58	69.96	32.40	16.51	16.13	1	1	1	32	17	16	9.5	6	6	9.5	6	6
54.06	86.18	68.60	69.81	32.12	17.58	16.37	1	1	1	32	18	16	9	13	5.5	9	13	5.5
53.83	86.37	68.40	69.49	32.53	17.96	16.87	1	1	1	33	18	17	19	12.5	15.5	19	12.5	15.5
54.02	86.32	68.41	69.96	32.29	17.91	16.35	1	1	1	32	18	16	8.5	12	5	8.5	12	5
53.46	85.70	67.43	69.26	32.24	18.27	16.44	1	1	1	32	18	16	8	11.5	4.5	8	11.5	4.5
53.82	86.10	68.38	69.41	32.28	17.72	16.69	1	1	1	32	18	17	7.5	11	13	7.5	11	13
54.46	86.48	68.75	69.88	32.02	17.73	16.60	1	1	1	32	18	17	7	10.5	12.5	7	10.5	12.5
53.81	86.25	69.29	69.95	32.45	16.97	16.31	1	1	1	32	17	16	6.5	5.5	4	6.5	5.5	4
53.97	86.24	68.55	70.01	32.26	17.69	16.22	1	1	1	32	18	16	6	9	3.5	6	9	3.5
53.53	86.18	71.06	69.47	32.66	15.12	16.72	1	1	1	33	15	17	12.5	1.5	10	12.5	1.5	10
53.87	86.44	67.83	69.67	32.57	18.61	16.77	1	1	1	33	19	17	12	12	9.5	12	12	9.5
53.86	86.01	67.97	69.63	32.15	18.04	16.38	1	1	1	32	18	16	5.5	7.5	3	5.5	7.5	3
54.19	86.11	70.77	69.54	31.92	15.34	16.57	1	1	1	32	15	17	5	1	8	5	1	8
53.89	86.12	68.40	69.95	32.23	17.72	16.17	1	1	1	32	18	16	4.5	6	2.5	4.5	6	2.5
53.90	86.07	67.39	69.47	32.17	18.69	16.60	1	1	1	32	19	17	4	8.5	6.5	4	8.5	6.5
53.83	86.18	67.32	69.39	32.35	18.86	16.78	1	1	1	32	19	17	3.5	8	6	3.5	8	6
53.80	86.53	70.52	69.66	32.73	16.01	16.86	1	1	1	33	16	17	6.5	1.5	5.5	6.5	1.5	5.5
54.01	86.07	68.70	69.80	32.06	17.36	16.26	1	1	1	32	17	16	3	2	2	3	2	2
53.81	86.04	67.73	69.63	32.23	18.30	16.41	1	1	1	32	18	16	2.5	3.5	1.5	2.5	3.5	1.5
53.67	86.22	68.57	69.53	32.55	17.66	16.70	1	1	1	33	18	17	4	3	3	4	3	3
54.03	86.46	70.78	69.82	32.43	15.68	16.65	1	1	1	32	16	17	2	1	2.5	2	1	2.5
54.00	86.35	68.82	69.70	32.35	17.53	16.65	1	1	1	32	18	17	1.5	1.5	2	1.5	1.5	2
53.91	86.18	67.91	69.71	32.27	18.28	16.48	1	1	1	32	18	16	1	1	1	1	1	1
Sum of negative ranks(W-)																**0**	**0**	**0**
Sum of positive ranks(W+)																**693**	**697**	**662.5**

Table H.2: Wilcoxon's sign rank test based on 10 times 5-fold cross validation. This result is based on dataset #2 using the motion feature set without feature selection.

Correct Classification				DIFFERENCE			SIGN			ABS			RANK			SiGN X RANK		
NB	RF	RT	SVM	RF-NB	RF-RT	RF-SVM	RF-NB	RF-RT	RF-SVM	|RF-NB|	|RF-RT|	|RF-SVM|	RF-NB	RF-RT	RF-SVM	RF-NB	RF-RT	RF-SVM
68.41	88.63	73.32	70.72	20.22	15.31	17.90	1	1	1	20	15	18	19	33.5	23	19	33.5	23
68.36	88.64	74.26	70.41	20.27	14.37	18.23	1	1	1	20	14	18	18.5	14.5	22.5	18.5	14.5	22.5
68.25	88.60	74.86	70.99	20.35	13.74	17.61	1	1	1	20	14	18	18	14	22	18	14	22
68.23	88.58	73.75	70.02	20.35	14.83	18.56	1	1	1	20	15	19	17.5	31	45	17.5	31	45
68.31	88.53	74.84	70.24	20.21	13.69	18.29	1	1	1	20	14	18	17	13.5	21.5	17	13.5	21.5
68.03	88.71	73.49	70.65	20.68	15.22	18.06	1	1	1	21	15	18	39	29.5	21	39	29.5	21
68.15	88.75	73.64	70.44	20.61	15.11	18.31	1	1	1	21	15	18	38.5	29	20.5	38.5	29	20.5
68.41	88.54	74.64	70.11	20.14	13.90	18.43	1	1	1	20	14	18	16.5	13	20	16.5	13	20
68.83	88.64	75.05	70.58	19.81	13.58	18.06	1	1	1	20	14	18	16	12.5	19.5	16	12.5	19.5
68.22	88.70	74.32	70.66	20.48	14.38	18.04	1	1	1	20	14	18	15.5	12	19	15.5	12	19
68.61	88.89	73.88	70.74	20.28	15.01	18.15	1	1	1	20	15	18	15	25.5	18.5	15	25.5	18.5
68.48	88.45	74.18	70.34	19.97	14.27	18.11	1	1	1	20	14	18	14.5	11.5	18	14.5	11.5	18
67.94	88.59	75.69	70.67	20.65	12.91	17.93	1	1	1	21	13	18	33	3.5	17.5	33	3.5	17.5
68.36	88.50	73.55	70.23	20.14	14.96	18.28	1	1	1	20	15	18	14	23	17	14	23	17
68.40	88.74	74.00	70.56	20.34	14.74	18.17	1	1	1	20	15	18	13.5	22.5	16.5	13.5	22.5	16.5
68.31	88.74	74.52	70.67	20.43	14.21	18.07	1	1	1	20	14	18	13	10	16	13	10	16
68.05	88.51	74.47	70.22	20.46	14.03	18.29	1	1	1	20	14	18	12.5	9.5	15.5	12.5	9.5	15.5
67.91	89.00	74.72	70.16	21.09	14.28	18.85	1	1	1	21	14	19	28.5	9	31.5	28.5	9	31.5
68.54	88.97	74.49	70.80	20.44	14.48	18.17	1	1	1	20	14	18	12	8.5	15	12	8.5	15
68.68	88.57	73.33	70.53	19.89	15.23	18.03	1	1	1	20	15	18	11.5	18	14.5	11.5	18	14.5
68.35	89.06	74.78	70.47	20.71	14.29	18.59	1	1	1	21	14	19	26	8	29	26	8	29
68.31	88.66	75.30	70.57	20.35	13.36	18.09	1	1	1	20	13	18	11	3	14	11	3	14
68.31	88.18	73.26	70.40	19.87	14.91	17.78	1	1	1	20	15	18	10.5	15.5	13.5	10.5	15.5	13.5
68.21	88.54	74.74	70.33	20.33	13.81	18.21	1	1	1	20	14	18	10	6.5	13	10	6.5	13
68.49	88.90	73.21	70.59	20.41	15.68	18.31	1	1	1	20	16	18	9.5	23.5	12.5	9.5	23.5	12.5
68.18	88.50	73.11	70.32	20.32	15.39	18.18	1	1	1	20	15	18	9	14	12	9	14	12
68.04	88.61	73.87	70.76	20.56	14.73	17.85	1	1	1	21	15	18	20.5	13.5	11.5	20.5	13.5	11.5
68.66	88.66	73.03	70.60	20.00	15.63	18.06	1	1	1	20	16	18	8.5	21	11	8.5	21	11
68.52	88.43	74.01	70.34	19.91	14.42	18.09	1	1	1	20	14	18	8	6	10.5	8	6	10.5
68.27	88.73	73.32	70.44	20.46	15.40	18.29	1	1	1	20	15	18	7.5	12	10	7.5	12	10
68.73	88.75	73.77	70.29	20.02	14.98	18.45	1	1	1	20	15	18	7	11.5	9.5	7	11.5	9.5
68.24	88.39	73.13	70.74	20.15	15.26	17.65	1	1	1	20	15	18	6.5	11	9	6.5	11	9
67.93	88.63	75.16	70.07	20.70	13.46	18.56	1	1	1	21	13	19	15	2.5	17.5	15	2.5	17.5
68.25	88.89	72.92	71.07	20.65	15.97	17.83	1	1	1	21	16	18	14.5	15.5	8.5	14.5	15.5	8.5
68.27	88.60	73.78	70.23	20.34	14.82	18.37	1	1	1	20	15	18	6	9.5	8	6	9.5	8
68.27	89.00	73.82	70.56	20.73	15.18	18.44	1	1	1	21	15	18	13	9	7.5	13	9	7.5
68.24	88.98	72.59	70.77	20.75	16.40	18.21	1	1	1	21	16	18	12.5	13	7	12.5	13	7
68.08	88.40	73.37	70.11	20.32	15.03	18.29	1	1	1	20	15	18	5.5	8.5	6.5	5.5	8.5	6.5
68.73	88.70	73.04	70.65	19.96	15.65	18.05	1	1	1	20	16	18	5	11.5	6	5	11.5	6
68.13	88.46	73.61	70.45	20.33	14.84	18.01	1	1	1	20	15	18	4.5	8	5.5	4.5	8	5.5
68.23	88.76	75.73	70.64	20.53	13.03	18.11	1	1	1	21	13	18	9	2	5	9	2	5
68.71	88.74	73.98	70.35	20.03	14.76	18.39	1	1	1	20	15	18	4	6.5	4.5	4	6.5	4.5
68.38	88.55	73.78	70.31	20.18	14.77	18.25	1	1	1	20	15	18	3.5	6	4	3.5	6	4
68.17	88.78	75.84	70.47	20.60	12.94	18.30	1	1	1	21	13	18	6.5	1.5	3.5	6.5	1.5	3.5
68.12	88.46	74.05	70.68	20.34	14.41	17.78	1	1	1	20	14	18	3	2.5	3	3	2.5	3
68.35	88.52	72.80	70.86	20.18	15.73	17.67	1	1	1	20	16	18	2.5	5	2.5	2.5	5	2.5
68.16	88.86	74.76	70.28	20.70	14.10	18.57	1	1	1	21	14	19	4	2	4	4	2	4
68.49	88.83	73.81	70.39	20.34	15.02	18.45	1	1	1	20	15	18	2	2.5	2	2	2.5	2
68.38	88.43	73.78	70.55	20.05	14.64	17.88	1	1	1	20	15	18	1.5	2	1.5	1.5	2	1.5
68.36	88.54	75.06	70.38	20.18	13.48	18.15	1	1	1	20	13	18	1	1	1	1	1	1
Sum of negative ranks (W-)																**0**	**0**	**0**
Sum of positive ranks (W+)																**630**	**597.5**	**667**

Table H.3: Wilcoxon's sign rank test based on 10 times 5-fold cross validation. This result is based on dataset #2 using the combined feature set without feature selection.

Correct Classification				DIFFERENCE			SIGN			ABS			RANK			SiGN X RANK		
NB	RF	RT	SVM	RF-NB	RF-RT	RF-SVM	RF-NB	RF-RT	RF-SVM	|RF-NB|	|RF-RT|	|RF-SVM|	RF-NB	RF-RT	RF-SVM	RF-NB	RF-RT	RF-SVM
69.60	84.87	65.83	75.89	15.27	19.04	8.97	1	1	1	15	19	9	13	14.5	21.5	13	14.5	21.5
69.39	85.28	67.41	75.74	15.88	17.87	9.54	1	1	1	16	18	10	37	6	45.5	37	6	45.5
69.38	85.02	65.41	75.92	15.63	19.61	9.10	1	1	1	16	20	9	36.5	26	21	36.5	26	21
69.29	84.90	66.10	75.76	15.60	18.80	9.14	1	1	1	16	19	9	36	13	20.5	36	13	20.5
69.55	84.90	65.46	75.61	15.35	19.44	9.29	1	1	1	15	19	9	12.5	12.5	20	12.5	12.5	20
69.11	84.82	65.05	75.80	15.72	19.78	9.02	1	1	1	16	20	9	34.5	23.5	19.5	34.5	23.5	19.5
69.42	84.95	66.82	75.81	15.53	18.13	9.14	1	1	1	16	18	9	34	5.5	19	34	5.5	19
69.14	84.55	67.52	75.26	15.41	17.03	9.28	1	1	1	15	17	9	12	2.5	18.5	12	2.5	18.5
70.02	85.10	65.35	75.98	15.08	19.75	9.11	1	1	1	15	20	9	11.5	21	18	11.5	21	18
69.32	85.24	64.46	75.98	15.93	20.78	9.26	1	1	1	16	21	9	31.5	31.5	17.5	31.5	31.5	17.5
69.62	84.70	64.00	76.15	15.09	20.70	8.55	1	1	1	15	21	9	11	31	17	11	31	17
68.98	84.60	64.99	75.41	15.61	19.61	9.19	1	1	1	16	20	9	30	20.5	16.5	30	20.5	16.5
69.29	85.01	64.76	75.84	15.72	20.25	9.17	1	1	1	16	20	9	29.5	20	16	29.5	20	16
69.43	84.99	65.42	75.52	15.57	19.57	9.47	1	1	1	16	20	9	29	19.5	15.5	29	19.5	15.5
69.34	84.72	66.22	75.91	15.37	18.50	8.81	1	1	1	15	18	9	10.5	4	15	10.5	4	15
69.76	85.40	66.77	75.70	15.64	18.63	9.69	1	1	1	16	19	10	27.5	9	32	27.5	9	32
68.73	85.20	65.81	75.64	16.47	19.38	9.55	1	1	1	16	19	10	27	8.5	31.5	27	8.5	31.5
69.40	84.87	66.70	75.66	15.46	18.16	9.21	1	1	1	15	18	9	10	3.5	14.5	10	3.5	14.5
69.57	84.66	64.76	76.00	15.09	19.91	8.67	1	1	1	15	20	9	9.5	15	14	9.5	15	14
69.65	84.69	63.01	75.84	15.04	21.68	8.85	1	1	1	15	22	9	9	28.5	13.5	9	28.5	13.5
69.48	84.54	64.02	75.85	15.06	20.53	8.69	1	1	1	15	21	9	8.5	22.5	13	8.5	22.5	13
69.54	84.92	63.20	75.86	15.37	21.71	9.06	1	1	1	15	22	9	8	27	12.5	8	27	12.5
69.27	84.93	63.62	75.85	15.66	21.30	9.07	1	1	1	16	21	9	21.5	22	12	21.5	22	12
69.40	84.61	63.95	75.65	15.20	20.66	8.96	1	1	1	15	21	9	7.5	21.5	11.5	7.5	21.5	11.5
69.35	84.88	62.63	75.80	15.53	22.25	9.09	1	1	1	16	22	9	20	24.5	11	20	24.5	11
69.29	84.65	65.94	75.83	15.36	18.71	8.82	1	1	1	15	19	9	7	7	10.5	7	7	10.5
69.48	85.38	66.83	75.54	15.90	18.55	9.84	1	1	1	16	19	10	18.5	6.5	22	18.5	6.5	22
69.80	84.66	69.08	76.02	14.86	15.58	8.65	1	1	1	15	16	9	6.5	1	10	6.5	1	10
69.34	84.63	65.54	75.69	15.29	19.10	8.94	1	1	1	15	19	9	6	5	9.5	6	5	9.5
69.11	85.39	66.68	75.64	16.28	18.71	9.75	1	1	1	16	19	10	16	4.5	19.5	16	4.5	19.5
69.66	85.07	63.81	75.73	15.41	21.25	9.33	1	1	1	15	21	9	5.5	16	9	5.5	16	9
69.89	85.24	64.33	76.14	15.36	20.92	9.11	1	1	1	15	21	9	5	15.5	8.5	5	15.5	8.5
69.04	84.53	63.91	75.49	15.49	20.62	9.04	1	1	1	15	21	9	4.5	15	8	4.5	15	8
68.96	84.99	66.43	75.98	16.04	18.57	9.01	1	1	1	16	19	9	12.5	4	7.5	12.5	4	7.5
69.44	84.34	64.19	75.56	14.91	20.15	8.79	1	1	1	15	20	9	4	8.5	7	4	8.5	7
69.20	84.82	64.75	75.70	15.62	20.07	9.12	1	1	1	16	20	9	11	8	6.5	11	8	6.5
69.45	84.88	65.37	75.78	15.43	19.51	9.10	1	1	1	15	20	9	3.5	7.5	6	3.5	7.5	6
69.21	83.88	65.53	75.09	14.67	18.35	8.79	1	1	1	15	18	9	3	2	5.5	3	2	5.5
69.50	85.20	65.03	76.06	15.69	20.16	9.13	1	1	1	16	20	9	8.5	6	5	8.5	6	5
69.75	85.18	63.80	76.00	15.43	21.38	9.18	1	1	1	15	21	9	2.5	8.5	4.5	2.5	8.5	4.5
69.48	84.77	63.28	75.73	15.29	21.49	9.03	1	1	1	15	21	9	2	8	4	2	8	4
69.38	85.13	68.31	75.92	15.74	16.81	9.21	1	1	1	16	17	9	6	1	3.5	6	1	3.5
69.18	85.20	65.07	75.58	16.02	20.12	9.62	1	1	1	16	20	10	5.5	4.5	7	5.5	4.5	7
69.30	85.30	65.03	75.63	16.00	20.27	9.67	1	1	1	16	20	10	5	4	6.5	5	4	6.5
69.35	85.16	65.91	75.87	15.81	19.25	9.29	1	1	1	16	19	9	4.5	1.5	3	4.5	1.5	3
69.34	85.17	61.79	75.88	15.84	23.39	9.29	1	1	1	16	23	9	4	5	2.5	4	5	2.5
69.63	85.39	65.77	75.63	15.76	19.62	9.75	1	1	1	16	20	10	3.5	2.5	4	3.5	2.5	4
69.29	84.55	66.01	75.80	15.27	18.54	8.76	1	1	1	15	19	9	1.5	1	2	1.5	1	2
69.64	84.88	65.03	75.77	15.24	19.85	9.11	1	1	1	15	20	9	1	1	1.5	1	1	1.5
69.21	85.35	63.17	75.92	16.14	22.17	9.42	1	1	1	16	22	9	1	1	1	1	1	1
Sum of negative ranks (W-)																**0**	**0**	**0**
Sum of positive ranks (W+)																**665**	**577.5**	**640.5**

References

Ainley, D. G., Porzig, E., Zajanc, D., and Spear, L. B. (2015). Seabird flight behaviour and height in response to altered wind strength and direction. *Marine Ornithology*, 43:25–36.

Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid, C. (2014). Good practice in large-scale learning for image classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(3):507–520.

Ammerman, L. K., McDonough, M., Hristov, N. I., and Kunz, T. H. (2009). Census of the endangered mexican long-nosed bat leptonycteris nivalis in texas, usa, using thermal imaging. *Endangered Species Research*, 8(1-2):87–92.

Anand, A., Pugalenthi, G., Fogel, G. B., and Suganthan, P. (2010). An approach for classification of highly imbalanced data using weighting and undersampling. *Amino acids*, 39(5):1385–1391.

Anwar, H., Zambanini, S., and Kampel, M. (2015a). Efficient scale-and rotation-invariant encoding of visual words for image classification. *Signal Processing Letters, IEEE*, 22(10):1762–1765.

Anwar, H., Zambanini, S., and Kampel, M. (2015b). Invariant image-based species classification of butterflies and reef fish. In T. Amaral, S. Matthews, T. P. S. M. and Fisher, R., editors, *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB)*, pages 5.1–5.8. BMVA Press.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188.

Ayyildiz, K. and Conrad, S. (2011). Video classification by main frequencies of repeating movements. In *12th Int. Workshop on Image Analysis for Multimedia Interactive Services*.

Baillie, S., Marchant, J., Leech, D., Massimino, D., Sullivan, M., Eglington, S., Barimore, C., Dadam, D., Downie, I., Harris, S., et al. (2014). Birdtrends 2014: trends in numbers, breeding success and survival for uk breeding birds.

Barandela, R., Valdovinos, R. M., and Sánchez, J. S. (2003). New applications of ensembles of classifiers. *Pattern Analysis & Applications*, 6(3):245–256.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., and Frommolt, K.-H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, 31(12):1524–1534.

Barua, S., Islam, M. M., Yao, X., and Murase, K. (2014). Mwmote–majority weighted minority oversampling technique for imbalanced data set learning. *Knowledge and Data Engineering, IEEE Transactions on*, 26(2):405–425.

Bashir, F. I., Khokhar, A. A., and Schonfeld, D. (2006). View-invariant motion trajectory-based activity classification and recognition. *Multimedia Systems*, 12(1):45–54.

Batista, G. E., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1):20–29.

Batuwita, R. and Palade, V. (2013). Class imbalance learning methods for support vector machines. *Imbalanced learning: Foundations, algorithms, and applications*, pages 83–99.

Berg, T. and Belhumeur, P. N. (2013). Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 955–962. IEEE.

Berg, T., Liu, J., Lee, S. W., Alexander, M. L., Jacobs, D. W., and Belhumeur, P. N. (2014). Birdsnap: Large-scale fine-grained visual categorization of birds. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2019–2026. IEEE.

Betke, M., Hirsh, D. E., Makris, N. C., McCracken, G. F., Procopio, M., Hristov, N. I., Tang, S., Bagchi, A., Reichard, J. D., Horn, J. W., et al. (2008). Thermal imaging reveals significantly smaller brazilian free-tailed bat colonies than previously estimated. *Journal of Mammalogy*, 89(1):18–24.

Beyan, Ç. (2015). Detection of unusual fish trajectories from underwater videos.

Beyan, C. and Fisher, R. B. (2013a). Detecting abnormal fish trajectories using clustered and labeled data. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 1476–1480. IEEE.

Beyan, C. and Fisher, R. B. (2013b). Detection of abnormal fish trajectories using a clustering based hierarchical classifier. In *British Machine Vision Conference (BMVC), Bristol, UK*.

Bhattacharya, U. and Chaudhuri, B. (2003). A majority voting scheme for multiresolution recognition of handprinted numerals. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pages 16–20. IEEE.

Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1):1063–1095.

Blair, R. B. (1999). Birds and butterflies along an urban gradient: surrogate taxa for assessing biodiversity? *Ecological applications*, 9(1):164–170.

Bouckaert, R. R. and Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–12. Springer.

Bouwmans, T., El Baf, F., and Vachon, B. (2008). Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3):219–237.

Branson, S., Van Horn, G., Belongie, S., and Perona, P. (2014). Bird species categorization using pose normalized deep convolutional nets. *arXiv preprint arXiv:1406.2952*.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breslav, M., Fuller, N. W., and Betke, M. (2012). Vision system for wingbeat analysis of

bats in the wild. In *Proceedings of the Workshop on Visual Observation and Analysis of Animal and Insect Behavior (VAIB 2012), held in conjunction with the 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba, Japan*. Citeseer.

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J., Hadley, A. S., and Betts, M. G. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6):4640–4650.

Bruderer, B., Peter, D., Boldt, A., and Liechti, F. (2010). Wing-beat characteristics of birds recorded with tracking radar and cine camera. *Ibis*, 152(2):272–291.

Buckland, S., Baillie, S., Dick, J. M., Elston, D., Magurran, A., Scott, E., Smith, R., Somerfield, P., Studeny, A., and Watt, A. (2012). How should regional biodiversity be monitored? *Environmental and Ecological Statistics*, 19(4):601–626.

Bullen, R. and McKenzie, N. (2002). Scaling bat wingbeat frequency and amplitude. *Journal of experimental biology*, 205(17):2615–2626.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

Cawley, G. C. and Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357.

Chawla, N. V., Lazarevic, A., Hall, L. O., and Bowyer, K. W. (2003). Smoteboost: Improving prediction of the minority class in boosting. In *Knowledge Discovery in Databases: PKDD 2003*, pages 107–119. Springer.

Chen, C.-h., Pau, L.-F., and Wang, P. S.-p. (2010). *Handbook of pattern recognition and computer vision*, volume 27. World Scientific.

Cochran, W. W., Bowlin, M. S., and Wikelski, M. (2008). Wingbeat frequency and flap-pause ratio during natural migratory flight in thrushes. *Integrative and comparative biology*, 48(1):134–151.

Corani, G. and Benavoli, A. (2015). A bayesian approach for comparing cross-validated algorithms on multiple data sets. *Machine Learning*, 100(2-3):285–304.

Cullinan, V. I., Matzner, S., and Duberstein, C. A. (2015). Classification of birds and bats using flight tracks. *Ecological Informatics*, 27:55–63.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11):2783–2792.

Cutler, R. and Davis, L. S. (2000). Robust real-time periodic motion detection, analysis, and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):781–796.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30.

Deori, B. and Thounaojam, D. M. (2014). A survey on moving object tracking in video. *International Journal on Information Theory (IJIT)*, 3(3):31–46.

Díaz-Uriarte, R. and De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1):3.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923.

Dietterich, T. G. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 2:263–286.

Domingos, P. and Pazzani, M. (1997). On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29(2-3):103–130.

Du, J.-X., Wang, X.-F., and Zhang, G.-J. (2007). Leaf shape based plant species recognition. *Applied mathematics and computation*, 185(2):883–893.

Duan, K., Parikh, D., Crandall, D., and Grauman, K. (2012). Discovering localized attributes for fine-grained recognition. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3474–3481. IEEE.

Duberstein, C., Virden, D., Matzner, S., Myers, J., Cullinan, V., and Maxwell, A. (2012). Automated thermal image processing for detection and classification of birds and bats.

Eaton, M., Aebischer, N., Brown, A., Hearn, R., Lock, L., Musgrove, A., Noble, D., Stroud, D., and Gregory, R. (2015). Birds of conservation concern 4: the population status of birds in the uk, channel islands and isle of man. *British Birds*, 108:708–746.

Evans, W. R. and Rosenberg, K. V. (2000). Acoustic monitoring of night-migrating birds: a progress report. *Strategies for bird conservation: The Partners in Flight planning process*, pages 1–5.

Fan, W., Wang, H., Yu, P. S., and Ma, S. (2003). Is random model better? on its accuracy and efficiency. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 51–58. IEEE.

Fay, R. (2007). *Feature selection and information fusion in hierarchical neural networks for iterative 3D-object recognition.* PhD thesis, University of Ulm.

Fisher, R., Boom, B., and Huang, P. Preliminary experiments with the fish4knowledge dataset. *Algae*, 49165(49370):99–58.

Fouad, M. M. M., Zawbaa, H. M., El-Bendary, N., and Hassanien, A. E. (2013). Automatic nile tilapia fish classification approach using machine learning techniques. In *Hybrid Intelligent Systems (HIS), 2013 13th International Conference on*, pages 173–178. IEEE.

Frank, E. and Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 503–510. Springer.

Freund, Y., Schapire, R. E., et al. (1996). Experiments with a new boosting algorithm. In *Icml*, volume 96, pages 148–156.

Frezza, H. (2013). Support vector machines tutorial.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*,

volume 1. Springer series in statistics Springer, Berlin.

García, S., Cano, J. R., Fernández, A., and Herrera, F. (2006). A proposal of evolutionary prototype selection for class imbalance problems. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 1415–1423. Springer.

García, S. and Herrera, F. (2009). Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. *Evolutionary computation*, 17(3):275–306.

García, V., Sánchez, J. S., and Mollineda, R. A. (2012). On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. *Knowledge-Based Systems*, 25(1):13–21.

Gavves, E., Fernando, B., Snoek, C. G., Smeulders, A. W., and Tuytelaars, T. (2013). Fine-grained categorization by alignments. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1713–1720. IEEE.

Gavves, E., Fernando, B., Snoek, C. G., Smeulders, A. W., and Tuytelaars, T. (2015). Local alignments for fine-grained categorization. *International Journal of Computer Vision*, 111(2):191–212.

Ge, Z., McCool, C., Sanderson, C., and Corke, P. (2015). Subset feature learning for fine-grained category classification. *arXiv preprint arXiv:1505.02269*.

Ghaderian, M., Behrad, A., and Kaboodi, S. A. D. (2011). Recognition of periodic motions using one-dimensional contour based features. In *Machine Vision and Image Processing, 2011 7th Iranian*, pages 1–5. IEEE.

Gonzalez, R. C. and Woods, R. E. (2002). Digital image processing.

Goodenough, A. E., Fairhurst, S. M., Morrison, J. B., Cade, M., Morgan, P. J., and Wood, M. J. (2014). Quantifying the robustness of first arrival dates as a measure of avian migratory phenology. *Ibis*.

Gouyon, F., Pachet, F., and Delerue, O. (2000). Classifying percussive sounds: a matter of zero-crossing rate. In *Proceedings of the COST G-6 Conference on Digital Audio Effects, Verona, Italy*, pages 7–9.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., and Schmidhuber,

J. (2009). A novel connectionist system for unconstrained handwriting recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(5):855–868.

Gregory, R. (2006). Birds as biodiversity indicators for europe. *Significance*, 3(3):106–110.

Gregory, R., Baillie, S., and Bashford, R. (2000). Monitoring breeding birds in the united kingdom. *Bird Census News*, 13:101–112.

Gregory, R. D., Gibbons, D. W., and Donald, P. F. (2004). Bird census and survey techniques. *Bird ecology and conservation*, pages 17–56.

Gregory, R. D., Gibbons, D. W., and Donald, P. F. (2015). Wild bird populations in the uk, 1970 to 2014.

Gu, Q., Li, Z., and Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Hald, A. (1999). On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science*, pages 214–222.

Hall, D. and Perona, P. (2015). Fine-grained classification of pedestrians in video: Benchmark and state of the art. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5482–5491.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

Hall, M., Witten, I., and Frank, E. (2011). Data mining: Practical machine learning tools and techniques. *Kaufmann, Burlington*.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato.

Hall, M. A. (2000). Correlation-based feature selection of discrete and numeric class machine learning.

Hammers, M., Müskens, G. J., van Kats, R. J., Teunissen, W. A., and Kleijn, D. (2014).

Ecological contrasts drive responses of wintering farmland birds to conservation management. *Ecography*.

Han, H., Wang, W.-Y., and Mao, B.-H. (2005). Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *Advances in intelligent computing*, pages 878–887. Springer.

Handa, A., Newcombe, R. A., Angeli, A., and Davison, A. J. (2012). Real-time camera tracking: When is high frame-rate best? In *European Conference on Computer Vision*, pages 222–235. Springer.

Harrison, P. J., Buckland, S. T., Yuan, Y., Elston, D. A., Brewer, M. J., Johnston, A., and Pearce-Higgins, J. W. (2014). Assessing trends in biodiversity over space and time using the example of british breeding birds. *Journal of Applied Ecology*, 51(6):1650–1660.

Hayhow, D., Conway, G., Eaton, M., Grice, P., Hall, C., Holt, C., Kuepfers, A., Noble, D., Oppel, S., Risely, K., et al. (2014). The state of the ukâĂŹs birds 2014. *RSPB, BTO, WWT, JNCC, NE, NIEA, NRW and SNH, Sandy, Bedfordshire*.

Hellerstein, J. L., Jayram, T., Rish, I., et al. (2000). *Recognizing end-user transactions in performance management*. IBM Thomas J. Watson Research Division.

Hristov, N. I., Betke, M., Theriault, D. E., Bagchi, A., and Kunz, T. H. (2010). Seasonal variation in colony size of brazilian free-tailed bats at carlsbad cavern based on thermal imaging. *Journal Information*, 91(1).

Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.

Hu, X.-G., Li, P.-p., Wu, X.-D., and Wu, G.-Q. (2007). A semi-random multiple decision-tree algorithm for mining data streams. *Journal of Computer Science and Technology*, 22(5):711–724.

Huang, C., Luo, B., Tang, L., Liu, Y., and Ma, J. (2013). Topic model based bird breed classification and annotation. In *Communications, Circuits and Systems (ICCCAS), 2013 International Conference on*, volume 2, pages 319–322. IEEE.

Huang, J. and Ling, C. X. (2005). Using auc and accuracy in evaluating learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on*, 17(3):299–310.

Huang, J., Lu, J., and Ling, C. X. (2003). Comparing naive bayes, decision trees, and svm with auc and accuracy. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pages 553–556. IEEE.

Huang, Z.-C., Chan, P. P., Ng, W. W., and Yeung, D. S. (2010). Content-based image retrieval using color moment and gabor texture feature. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, volume 2, pages 719–724. IEEE.

Jacob, M., Blu, T., and Unser, M. (2001). An exact method for computing the area moments of wavelet and spline curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):633–642.

Jagannathan, G., Pillaipakkamnatt, K., and Wright, R. N. (2009). A practical differentially private random decision tree classifier. In *2009 IEEE International Conference on Data Mining Workshops*, pages 114–121. IEEE.

Jager, J., Simon, M., Denzler, J., Wolff, V., Fricke-Neuderth, K., and Kruschel, C. (2015). Croatian fish dataset: Fine-grained classification of fish species in their natural habitat. In T. Amaral, S. Matthews, T. P. S. M. and Fisher, R., editors, *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB)*, pages 6.1–6.7. BMVA Press.

James, G. (1998). *Majority vote classifiers: theory and applications*. PhD thesis, Stanford University.

Johnston, A., Cook, A. S., Wright, L. J., Humphreys, E. M., and Burton, N. H. (2014). Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines. *Journal of Applied Ecology*, 51(1):31–41.

Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.-P., Planque, R., Rauber, A., Fisher, R., and Müller, H. (2014). Lifeclef 2014: multimedia life species identification challenges. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, pages 229–249. Springer.

Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.-P., Planqué,

R., Rauber, A., Palazzo, S., Fisher, B., et al. (2015). Lifeclef 2015: multimedia life species identification challenges. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 462–483. Springer.

KaewTraKulPong, P. and Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-based surveillance systems*, pages 135–144. Springer.

Karim, M. and Rahman, R. M. (2013). Decision tree and naïve bayes algorithm for classification and generation of actionable knowledge for direct marketing.

Keerthi, S. S., Chapelle, O., and DeCoste, D. (2006). Building support vector machines with reduced classifier complexity. *The Journal of Machine Learning Research*, 7:1493–1515.

Kilham, L. (1990). *The American crow and the common raven*. Number 10. Texas A&M University Press.

Kira, K. and Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In *AAAI*, volume 2, pages 129–134.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of relief. In *European conference on machine learning*, pages 171–182. Springer.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., et al. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30(1):25–36.

Krause, J., Jin, H., Yang, J., and Fei-Fei, L. (2015). Fine-grained recognition without part annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5546–5555.

Kubat, M., Matwin, S., et al. (1997). Addressing the curse of imbalanced training sets: one-sided selection. In *ICML*, volume 97, pages 179–186. Nashville, USA.

Laptev, I., Belongie, S. J., Perez, P., and Wills, J. (2005). Periodic motion detection and segmentation via approximate sequence alignment. In *Computer Vision. ICCV. Tenth IEEE International Conference on*, volume 1, pages 816–823. IEEE.

Lazarevic, L., Harrison, D., Southee, D., Wade, M., and Osmond, J. (2008). Wind farm

and fauna interaction: detecting bird and bat wing beats through cyclic motion analysis. *Int. Journal of Sustainable Engineering*, 1(1):60–68.

Lazebnik, S., Schmid, C., and Ponce, J. (2004). Semi-local affine parts for object recognition. In *British Machine Vision Conference (BMVC'04)*, pages 779–788. The British Machine Vision Association (BMVA).

Lee, D., Redd, S., Schoenberger, R., Xu, X., and Zhan, P. E. (2003). An automated fish species classification and migration monitoring system. In *Industrial Electronics Society, 2003. IECON'03. The 29th Annual Conference of the IEEE*, volume 2, pages 1080–1085. IEEE.

Lee, J. W., Lee, J. B., Park, M., and Song, S. H. (2005). An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics & Data Analysis*, 48(4):869–885.

Lee, M. S., Rhee, J.-K., Kim, B.-H., and Zhang, B.-T. (2009). Aesnb: active example selection with naïve bayes classifier for learning from imbalanced biomedical data. In *Bioinformatics and BioEngineering, 2009. BIBE'09. Ninth IEEE International Conference on*, pages 15–21. IEEE.

Lee, S., Park, Y.-T., dâĂŹAuriol, B. J., et al. (2012). A novel feature selection method based on normalized mutual information. *Applied Intelligence*, 37(1):100–120.

Lee, T. S. (1996). Image representation using 2d gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959–971.

Li, T., Zhang, C., and Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15):2429–2437.

Li, X., Hu, W., and Hu, W. (2006). A coarse-to-fine strategy for vehicle motion trajectory clustering. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 591–594. IEEE.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R news*, 2(3):18–22.

Lichman, M. (2013). Uci machine learning repository.

Liechti, F. and Bruderer, L. (2002). Wingbeat frequency of barn swallows and house martins: a comparison between free flight and wind tunnel experiments. *Journal of Experimental Biology*, 205(16):2461–2467.

Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., Cao, L., and Huang, T. (2011). Large-scale image classification: fast feature extraction and svm training. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1689–1696. IEEE.

Liu, H. and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4):491–502.

Liu, X., Wang, X., and Su, Q. (2015). Feature selection of medical data sets based on rs-relieff. In *Service Systems and Service Management (ICSSSM), 2015 12th International Conference on*, pages 1–5. IEEE.

Liu, X.-Y., Wu, J., and Zhou, Z.-H. (2009). Exploratory undersampling for class-imbalance learning. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(2):539–550.

Liwicki, M., Bunke, H., et al. (2006). Hmm-based on-line recognition of handwritten whiteboard notes. In *Tenth International Workshop on Frontiers in Handwriting Recognition*.

Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A., Silla Jr, C. N., and Koerich, A. L. (2011). Automatic bird species identification for large number of species. In *Multimedia (ISM), 2011 IEEE International Symposium on*, pages 117–122. IEEE.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems*, pages 431–439.

Lyons, R. (2010). How to interpolate in the time-domain by zero-padding in the frequency domain.

Mai, F., Chang, C., and Hung, Y. (2010). Affine-invariant shape matching and recognition under partial occlusion. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4605–4608. IEEE.

Malik, F. and Baharudin, B. (2013). The statistical quantized histogram texture features analysis for image retrieval based on median and laplacian filters in the dct domain. *The International Arab Journal of Information Technology*, 10(6):1–9.

Marini, A., Facon, J., and Koerich, A. L. (2013). Bird species classification based on color features. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 4336–4341. IEEE.

Marini, A., Turatti, A., Britto, A., and Koerich, A. (2015). Visual and acoustic identification of bird species. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2309–2313. IEEE.

Martín, J. A., Santos, M., and de Lope, J. (2010). Orthogonal variant moments features in image analysis. *Information Sciences*, 180(6):846–860.

Martínez, A. M. and Kak, A. C. (2001). Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228–233.

Mastriani, M. (2014). Rule of three for superresolution of still images with applications to compression and denoising. *arXiv preprint arXiv:1405.0632*.

Matzner, S., Cullinan, V. I., and Duberstein, C. A. (2015). Two-dimensional thermal video analysis of offshore bird and bat flight. *Ecological Informatics*, 30:20–28.

Mehta, M., Agrawal, R., and Rissanen, J. (1996). Sliq: A fast scalable classifier for data mining. In *International Conference on Extending Database Technology*, pages 18–32. Springer.

Meng, H., Pears, N., and Bailey, C. (2007). A human action recognition system for embedded computer vision application. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–6. IEEE.

Mikusiński, G., Gromadzki, M., and Chylarecki, P. (2001). Woodpeckers as indicators of forest bird diversity. *Conservation biology*, 15(1):208–217.

Mladenić, D. (2006). *Feature selection for dimensionality reduction*. Springer.

Moeslund, T. B. (2012). *Introduction to video and image processing: Building real systems and applications*. Springer Science & Business Media.

Mokhtarian, F., Abbasi, S., and Kittler, J. (1996). Robust and e cient shape indexing

through curvature scale space. In *Proceedings of the 1996 British Machine and Vision Conference BMVC*, volume 96.

Mollineda, R., Alejo, R., and Sotoca, J. (2007). The class imbalance problem in pattern classification and learning. In *II Congreso Español de Informática (CEDI 2007). ISBN*, pages 978–84.

Moore, J. H. and White, B. C. (2007). Tuning relieff for genome-wide genetic analysis. In *Evolutionary computation, machine learning and data mining in bioinformatics*, pages 166–175. Springer.

Nadeau, C. and Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, 52(3):239–281.

Nakariyakul, S. and Casasent, D. P. (2009). An improvement on floating search algorithms for feature subset selection. *Pattern Recognition*, 42(9):1932–1940.

Neal, L., Briggs, F., Raich, R., and Fern, X. Z. (2011). Time-frequency segmentation of bird song in noisy acoustic environments. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 2012–2015. IEEE.

Norberg, U. M. L. and Norberg, R. Å. (2012). Scaling of wingbeat frequency with body mass in bats and limits to maximum bat size. *The Journal of Exp. Bio.*, 215(5):711–722.

Oberg, J., Eguro, K., Bittner, R., and Forin, A. (2012). Random decision tree body part recognition using fpgas. In *22nd International Conference on Field Programmable Logic and Applications (FPL)*, pages 330–337. IEEE.

Oh, S., Lee, M. S., and Zhang, B.-T. (2011). Ensemble learning with active example selection for imbalanced biomedical data classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(2):316–325.

Orriols-Puig, A. and Bernadó-Mansilla, E. (2009). Evolutionary rule-based systems for imbalanced data sets. *Soft Computing*, 13(3):213–225.

Ou, J., Bai, X.-B., Pei, Y., Ma, L., and Liu, W. (2010). Automatic facial expression recognition using gabor filter and expression analysis. In *Computer Modeling and Simulation, 2010. ICCMS'10. Second International Conference on*, volume 2, pages 215–218. IEEE.

Parvin, H., Mohammadi, M., and Rezaei, Z. (2012). Face identification based on gabor-wavelet features. *JDCTA: International Journal of Digital Content Technology and its Applications*, 6(1):247–255.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238.

Peters, J., Ramanna, S., and Szczuka, M. (2003). Towards a line-crawling robot obstacle classification system: A rough set approach. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 575–575.

Plotnik, A. M. and Rock, S. M. (2002). Quantification of cyclic motion of marine animals from computer vision. In *OCEANS MTS/IEEE*, volume 3, pages 1575–1581. IEEE.

Podulka, S., Rohrbaugh, R. W., Bonney, R., et al. (2004). *Handbook of bird biology*. Cornell Lab of Ornithology Ithaca, New York.

Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125.

Pun, C.-M. and Lee, M.-C. (2003). Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):590–603.

Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning*, pages 307–323. Springer.

Ramyachitr, D. D. and Manikandan, P. (2014). Imbalanced dataset classification and solutions: A review. *International Journal Of Computing And Business Research (Ijcbr)*, 5(4).

Ren, Y., Fan, B., Lin, W., Yang, X., Li, H., Li, W., and Liu, D. (2011). An efficient framework for analyzing periodical activities in sports videos. In *Image and Signal Processing, 2011 4th Int. Congress on*, volume 1, pages 502–506. IEEE.

Rennie, J. D., Shih, L., Teevan, J., Karger, D. R., et al. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *ICML*, volume 3, pages 616–623. Washington DC).

Robinson, R., Marchant, J., Leech, D., Massimino, D., Sullivan, M., Eglington, S., Barimore, C., Dadam, D., Downie, I., Hammond, M., Harris, S., Noble, D., Walker, R., and Baillie, S. (2015). Birdtrends 2015: trends in numbers, breeding success and survival for uk breeding birds.

Robnik-Šikonja, M. and Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53(1-2):23–69.

Rodrigues, M. T., Pádua, F. L., Gomes, R. M., and Soares, G. E. (2010). Automatic fish species classification based on robust feature extraction techniques and artificial immune systems. In *Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010 IEEE Fifth International Conference on*, pages 1518–1525. IEEE.

Rosenhahn, B., Kersting, U., Andrew, S., Brox, T., Klette, R., and Seidel, H.-P. (2005). A silhouette based human motion tracking system. Technical report, CITR, The University of Auckland, New Zealand.

Rother, C., Kolmogorov, V., and Blake, A. (2004a). Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM.

Rother, C., Kolmogorov, V., and Blake, A. (2004b). "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.

Schenk, J., Kaiser, M., and Rigoll, G. (2009). Selecting features in on-line handwritten whiteboard note recognition: Sfs or sffs? In *2009 10th International Conference on Document Analysis and Recognition*, pages 1251–1254. IEEE.

Schiffner, I. and Srinivasan, M. V. (2016). Budgerigar flight in a varying environment: flight at distinct speeds? *Biology Letters*, 12(6):20160221.

Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of computer vision*, 37(2):151–172.

Schüldt, C., Laptev, I., and Caputo, B. (2004). Recognizing human actions: a local svm approach. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 32–36. IEEE.

Schumacher, D. (1992). General filtered image rescaling. In *Graphics Gems III*, pages 8–16. Academic Press Professional, Inc.

Sebe, N., Lew, M. S., Cohen, I., Garg, A., and Huang, T. S. (2002). Emotion recognition using a cauchy naive bayes classifier. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 17–20. IEEE.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. (2010). Rusboost: A hybrid approach to alleviating class imbalance. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(1):185–197.

Sergyan, S. (2008). Color histogram features based image classification in content-based image retrieval systems. In *Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on*, pages 221–224. IEEE.

Shardlow, M. (2016). An analysis of feature selection techniques. *The University of Manchester*.

Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE.

Silla, C. N., Kaestner, C., et al. (2013). Hierarchical classification of bird species using their audio recorded songs. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 1895–1900. IEEE.

Sobran, N. M. M., Ahmad, A., and Ibrahim, Z. (2013). Classification of imbalanced dataset using conventional naive bayes classifier. In *International Conference on Artificial Intelligence in Computer Science and ICT*, pages 35–42.

Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.-H. J., Fisher, R. B., and Nadarajan, G. (2010). Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 45–50. ACM.

Spampinato, C., Palazzo, S., Boom, B., van Ossenbruggen, J., Kavasidis, I., Di Salvo, R., Lin, F.-P., Giordano, D., Hardman, L., and Fisher, R. B. (2014). Understanding fish behavior during typhoon events in real-life underwater environments. *Multimedia Tools and Applications*, 70(1):199–236.

Statnikov, A., Wang, L., and Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC bioinformatics*, 9(1):1.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE.

Suzuki, S. et al. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46.

Tan, L. N., Alwan, A., Kossan, G., Cody, M. L., and Taylor, C. E. (2015). Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training dataa). *The Journal of the Acoustical Society of America*, 137(3):1069–1080.

Tang, J., Alelyani, S., and Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, page 37.

Tang, Y., Zhang, Y.-Q., Chawla, N. V., and Krasser, S. (2009). Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288.

Tax, D. M. and Duin, R. P. (2002). Using two-class classifiers for multiclass classification. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 124–127. IEEE.

Tian, J., Satpathy, A., Ng, E. S., Ong, S. G., Cheng, W., Burgunder, J.-M., and Hunziker, W. (2014). Motion analytics of zebrafish using fine motor kinematics and multi-view trajectory. *Multimedia Systems*, pages 1–11.

Tobalske, B. W., Hearn, J. W., and Warrick, D. R. (2009). Aerodynamics of intermittent bounds in flying birds. *Experiments in fluids*, 46(5):963–973.

Vapnik, V. N. and Vapnik, V. (1998). *Statistical learning theory*, volume 1. Wiley New York.

Wah, C., Branson, S., Perona, P., and Belongie, S. (2011a). Multiclass recognition and part localization with humans in the loop. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2524–2531. IEEE.

Wah, C., Branson, S., Welinder, P., Perona, P., and Belongie, S. (2011b). The caltech-ucsd birds-200-2011 dataset.

Wang, J., Markert, K., and Everingham, M. (2009). Learning models for object recognition from natural language descriptions. In *BMVC*, volume 1, page 2.

Wang, N. X. R., Cullis-Suzuki, S., and Branzan Albu, A. (2015). Automated analysis of wild fish behavior in a natural habitat. In *Proceedings of the 2nd International Workshop on Environmental Multimedia Retrieval*, pages 21–26. ACM.

Wang, S. and Yao, X. (2009). Diversity analysis on imbalanced data sets by using ensemble models. In *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on*, pages 324–331. IEEE.

Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., and Perona, P. (2010). Caltech-ucsd birds 200.

Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. P. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):780–785.

Wu, M. and Wang, Y. (2015). A feature selection algorithm of music genre classification based on relieff and sfs. In *Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on*, pages 539–544. IEEE.

Yang, J. and Yang, J.-y. (2003). Why can lda be performed in pca transformed space? *Pattern recognition*, 36(2):563–566.

Yao, D., Yang, J., and Zhan, X. (2013). An improved random forest algorithm for class-imbalanced data classification and its application in pad risk factors analysis. *Open Electrical & Electronic Engineering Journal*, 7(1):62–70.

Yap, B. W., Rani, K. A., Rahman, H. A. A., Fong, S., Khairudin, Z., and Abdullah,

N. N. (2014). An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, pages 13–22. Springer.

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13.

Yilmaz, A., Li, X., and Shah, M. (2004). Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11):1531–1536.

Yokoyama, M. and Poggio, T. (2005). A contour-based moving object detection and tracking. In *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 271–276. IEEE.

Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML*, volume 3, pages 856–863.

Zaugg, S., Saporta, G., Van Loon, E., Schmaljohann, H., and Liechti, F. (2008). Automatic identification of bird targets with radar via patterns produced by wing flapping. *Journal of the Royal Society interface*, 5(26):1041–1053.

Zhang, H. (2004). The optimality of naive bayes. *AA*, 1(2):3.

Zhang, H. and Su, J. (2008). Naive bayes for optimal ranking. *Journal of Experimental & Theoretical Artificial Intelligence*, 20(2):79–93.

Zhang, X., Yuan, Q., Zhao, S., Fan, W., Zheng, W., and Wang, Z. (2010). Multi-label classification without the multi-label cost. In *SDM*, volume 10, pages 778–789. SIAM.

Zhang, Y., Fu, P., Liu, W., and Chen, G. (2014). Imbalanced data classification based on scaling kernel-based support vector machine. *Neural Computing and Applications*, 25(3-4):927–935.

Zhang, Y., Wei, X.-s., Wu, J., Cai, J., Lu, J., Nguyen, V.-A., and Do, M. N. (2015). Weakly supervised fine-grained image categorization. *arXiv preprint arXiv:1504.04943*.

Zheng, F. and Webb, G. I. (2005). A comparative study of semi-naive bayes methods in classification learning. In *Proceedings of the fourth Australasian data mining conference (AusDM05)*, pages 141–156. Citeseer.

Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE.

Zivkovic, Z. and van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780.